Nonstationary Continuous-Time Processes*

Federico M. Bandi
Graduate School of Business,
The University of Chicago

Peter C.B. Phillips
Cowles Foundation for Research in Economics,
Yale University

First version: September, 2001
This version: May 14, 2002

*__Preliminary - Comments are welcome__. Paper written for the Handbook of Financial Econometrics edited by Yacine Aït-Sahalia and Lars Peter Hansen. We thank Darrell Duffie, Benoit Perron and Mark Watson for discussions and Seoyeon Lee for research assistance. Bandi acknowledges financial support from the IBM Corporation Faculty Research Fund at the University of Chicago. Phillips thanks fhe NSF for support under grant no. SES 0092509. Comments are welcome and can be sent to Federico M. Bandi at federico.bandi@gsb.uchicago.edu.

Contents

1 Introduction

A large body of recent asset pricing theory is written in continuous-time, for which Merton (1990) and Duffie (1996) are classic references. Notwithstanding the evident benefit of continuous-time tools for modelling purposes and recent advances in the econometric treatment of continuous-time models,[1] the use of stochastic processes with continuous (in time) sample paths still poses remarkable challenges when it comes to the econometric estimation and empirical implementation of modern asset pricing models. Campbell, Lo and McKinlay (1997, CKL henceforth) and Gourieroux and Jasiak (2001) are recent textbooks on the general topic of financial econometrics and outline some of the relevant issues.

Perhaps the most basic econometric problem arises because, while the relevant series are often specified as processes that evolve continuously in time, observations of the process occur only at discrete points in time. The discrete nature of the data has forced researchers to design estimation methodologies that are capable of circumventing the so-called "aliasing problem" and that can uniquely identify the fine grain structure of the underlying process from a sample of observations located along the continuous sample path rather than from a continuous record of the process over that path. (Readers are referred to the Chapters by Aït-Sahalia *et al.* (2001), Bibby *et al.* (2001), Gallant and Tauchen (2001), Jacod (2001) and Johannes and Polson (2001) in the present volume for a treatment of these issues). Such methodologies generally, but not exclusively (c.f. Phillips (1973) and Hansen and Sargent (1983)), rely on stationarity. The reason is clear. Should the underlying process be endowed with a time-invariant stationary density, then the information extracted from the discrete data can fruitfully be employed to identify the time-invariant probability measure and thereby, hopefully, characterize the continuous dynamics of the system. In this way, stationarity can be a powerful aid to identification and estimation.

Despite the advantages of assuming the existence of a time-invariant stationary distribution, it appears that for many empirical applications in continuous-time asset pricing it would be more appropriate to allow for martingale and other forms of nonstationary behavior, while not ruling out stationarity either (see Section 6). In such cases, an additional layer of complication in estimation comes from the necessity of achieving identification without resorting to the restrictions that are provided by the existence of a time-invariant density for the process of interest.

This Chapter discusses techniques that have been recently introduced to identify potentially nonstationary, time-homogeneous, continuous-time Markov processes. The focus will be on classes of processes that are widely used in continuous-time asset pricing, namely scalar and multivariate diffusion processes as well as scalar jump-diffusion processes.

[1] In his survey on continuous-time methods in finance appeared in the Papers and Proceedings of the Sixtieth Annual Meeting of the American Finance Association, Sundaresan (2000) writes "Perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous-time."

Such processes, irrespective of their stationarity properties, have infinitesimal conditional moment definitions. Their infinitesimal moments are known to fully characterize the temporal evolution of the system and, in consequence, readily lend themselves to estimation for the purpose of the identification of the system's dynamics. Consider a standard scalar diffusion (i.e., the solution to (14) below), but a similar argument holds for more involved continuous-time Markov processes of the type reviewed in this Chapter. Its transition density (which is, in general, not known in closed-form) is fully determined by the two functions that are commonly known as the drift, $\mu(.)$, and the diffusion, $\sigma^2(.)$. The drift represents the conditional expected rate of change of the process for infinitesimal time changes, i.e.,

$$\mu(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}[X_t - X_0 | X_0 = a] = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a[X_t - a], \tag{1}$$

while the diffusion gives the conditional rate of change of volatility for infinitesimal variations in time, i.e.,

$$\sigma^2(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}[(X_t - X_0)^2 | X_0 = a] = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a[(X_t - a)^2]. \tag{2}$$

Formulae (1) and (2) are suggestive in that one could hope to identify the functions of interests, which are defined as conditional expectations over infinitesimal time distances, using sample analogues to conditional expectations as in standard nonparametric inference for conditional moments in discrete-time. For example, it is natural to estimate the drift at a by differencing the data and then averaging the first differences $X_{t+\Delta} - X_t$ corresponding to observations X_t's in the spatial neighborhood of the generic level a. Provided the level a is visited an infinite number of times over time so that an infinite number of differences can be averaged asymptotically, we would expect the procedure to be consistent in the limit (i.e., the sample average converges, at least in probability, to the conditional moment). Interestingly, the underlying process (and, under some conditions, the sampled process) visits the level a an infinite number of times provided recurrence is satisfied. Recurrence implies return of the sample path of the underlying process to any spatial set of non-zero Lebesgue measure with probability one and is known to be a milder assumption than stationarity (Section 2 provides a definition and additional discussion).

Some recent papers have pursued the econometric implications of these observations and designed spatial estimation methods for various classes of continuous-time, time-homogeneous, recurrent Markov processes. The methods are easy to implement and have some natural appeal because they are based on commonly used nonparametric (and semiparametric) estimation procedures for conditional moments in more conventional stationary, discrete-time, frameworks. Also, they have the additional attraction that their statistical properties apply even though stationarity of the underlying continuous-time model is never assumed. These methods have been developed in research by Bandi and Phillips (1998, BP (1998) hereafter), Bandi and Nguyen (2000, BN (2000) hereafter), and Bandi and Moloche (2001, BM (2001) hereafter) that develops kernel estimation procedures for scalar diffusions, scalar jump-diffusions and multivariate diffusions, respectively, and subsequently extended in various directions.

Specifically, following work by Brugiére (1991, 1993), Florens-Zmirou (1993) and Jacod (1997) in nonparametric diffusion function estimation for diffusions, this literature lays theoretical foundations for using well-understood and conventional nonparametric and semiparametric methods in the estimation of the infinitesimal moment functionals that drive the evolution of continuous-time Markov processes. The literature explores conditions (like recurrence) under which consistency and weak convergence results can be obtained in the (potential) absence of a stationary distribution for the process, and it provides results that can be evaluated in a manner that is closely related to conventional interpretations of nonparametric estimates for stationary discrete-time series. While the findings that emerge from this literature contain the stationary case as a subcase, their form in the more general case reflects the fact that a stationary density of the underlying process may not exist, which leads to important issues of interpretation.

This Chapter of the Handbook reviews these approaches and relates them to methods that are now well established in the nonparametric and semiparametric literature for discrete-time series. The Chapter also indicates some avenues for future research both in the estimation of potentially nonstationary continuous-time processes and in the estimation of potentially nonstationary discrete-time series.

The Chapter is organized as follows. Section 2 provides some intuition for the methodology, introduces the notion of recurrence that is used for identification in the potential absence of a stationary probability measure for the process of interest, and discusses the asymptotic features of the sampling scheme. In particular, consistency is shown to hinge on the joint implementation of "infill" and "long span" asymptotics. The latter is crucial in exploiting the recurrence properties of the process under investigation. The former is vital in replicating the infinitesimal features of the functions of interest. Both conditions are necessary for the identification of continuous-time Markov processes under minimal assumptions on their dynamic properties and parametric form. Accordingly, the discussion in this Chapter focuses on estimation procedures that impose mild conditions on the stochastic nature of the underlying process but require the presence, at least in the limit, of high frequency observations to achieve consistent estimation. In this regard, the present review is complementary to the Chapters in this Handbook by Aït-Sahalia *et al.* (2001) and by Johannes and Polson (2001). The former discusses functional estimation methods for diffusions that do not require infill asymptotics but which generally rely on stationarity and mixing for identification. The latter reviews Bayesian simulation procedures that are sufficiently flexible to deal with nonstationarities but impose a tight parametric structure on the process of interest.

Sections 3, 4 and 5 specialize the analysis to the estimation of recurrent scalar diffusions, recurrent jump-diffusions and recurrent multivariate diffusions, respectively. In Section 6 we provide an empirical example and discuss one of the approaches reviewed in this Chapter from a more applied perspective.

The Chapter is largely self-contained and its discussion is kept at a fairly intuitive level. Nonetheless, some basic notions of stochastic process theory and functional estimation in discrete-time econometrics will help the reader. Karatzas and Shreve (1988), Protter (1995) and Revuz and Yor (1998) are standard references for the former, and, while not providing all the background material for the present Chapter, they are strongly recommended references. Thorough discussions of functional methods for discrete-time series are contained in Härdle (1990) and Pagan and Ullah (1999). A more concise and highly accessible treatment of the same material is Härdle and Linton (1994). Chapter 12 of the book by CKL (1997) also provides an accessible discussion of kernel regression methods similar to those employed here.

2 Intuition and Conditions[2]

As noted in the Introduction, the existence of conditional moments for interesting classes of continuous-time Markov models provides a mechanism for inference based on the construction of sample analogues (i.e., weighted averages) to the infinitesimal conditional expectations. To fix ideas, take a simple example in discrete time. Suppose the observations $X_1, X_2 ... X_n$ are generated by a time-homogeneous Markov process X. One might be interested in estimating the conditional moment functional

$$\mathbf{M}(a) = \mathbf{E}^a[\mathbf{f}(X_1, a)], \tag{3}$$

where $X_0 = a$ is a generic initial condition and $\mathbf{f}$ is some integrable function. A crude (but intuitively appealing) sample analogue estimator for $\mathbf{M}(a)$ is

$$\widehat{\mathbf{M}}_{(n)}(a) = \frac{\sum_{i=1}^n \mathbf{1}_{X_i=a}\mathbf{f}(X_{i+1}, X_i)}{\sum_{i=1}^n \mathbf{1}_{X_i=a}}, \tag{4}$$

where $\mathbf{1}_A$ is the indicator function of the set A. Formula (4) implies identification of the conditional expectation (at $X_0 = a$) of the function $\mathbf{f}(X_1, X_0)$ (as in the definition of $\mathbf{M}(a)$) through a (weighted) sample average of functions of the observations $\mathbf{f}(X_{i+1}, X_i)$ taken at values X_i that are equal to a. Simple intuitive arguments based on the law of large numbers suggest that we need to visit the level a an infinite number of times to achieve consistency. In consequence, it appears that the condition $\#\{i : X_i = a\} = \sum_{i=1}^n \mathbf{1}_{X_i=a} \to \infty$ as $n \to \infty$ is, in general, necessary to obtain asymptotic convergence of $\widehat{\mathbf{M}}_{(n)}(a)$ to $\mathbf{M}(a)$.

[2]Parts of this section are based on the discussion of the paper "On the functional estimation of jump-diffusion processes" (BN (2000)) that Durrell Duffie gave at the 2001 Winter Meetings of the Econometric Society (New Orleans, January 9, 2001).

We now turn to a similar example in the context of a continuous-time Markov process X not necessarily endowed with a continuous sample path (one such case will be covered in Section 4 below). Suppose we are interested in estimating the infinitesimal conditional expectation

$$\mathbf{M}(a) = \lim_{t \to 0} \frac{1}{t} \mathbf{E}[\mathbf{f}(X_t, X_0) | X_0 = a] = \lim_{t \to 0} \frac{1}{t} \mathbf{E}^a[\mathbf{f}(X_t, a)]. \tag{5}$$

Note that if X is a scalar diffusion and $\mathbf{f}(y, a)$ is equal to either $(y - a)$ or $(y - a)^2$, then $\mathbf{M}(a)$ coincides with either the drift (1) or the diffusion function (2), respectively. Then, coherently with (4) above and the previous discussion, one could estimate (5) using

$$\widehat{\mathbf{M}}^{(1)}_{(n,\Delta,\varepsilon)}(a) = \frac{\sum_{i=1}^{n} \mathbf{1}_{X_i \in (a-\varepsilon, a+\varepsilon)} \mathbf{f}(X_{i+\Delta}, X_i)/\Delta}{\sum_{i=1}^{n} \mathbf{1}_{X_i \in (a-\varepsilon, a+\varepsilon)}}, \tag{6}$$

where Δ is the distance between observations and ε is a bandwidth parameter according to which we determine an interval around a on the sample path of the process. Asymptotically, we let Δ go to 0 to replicate the limit operation in the definition of $\mathbf{M}(a)$ (i.e. $\lim_{t \to 0}$). Furthermore, the bandwidth ε vanishes to zero in order to obtain averages of functions $\mathbf{f}(X_{i+\Delta}, X_i)$ such that X_i is in a close neighborhood of a and n grows to infinity to guarantee that the number of observations X_i in the actual vicinity of a (i.e., $\#\{i : X_i \in (a - \varepsilon, a + \varepsilon)\} = \sum_{i=1}^{n} \mathbf{1}_{X_i \in (a-\varepsilon, a+\varepsilon)}$) diverges to infinity for identification. Again, we expect $\widehat{\mathbf{M}}^{(1)}_{(n,\Delta,\varepsilon)}(a)$ to converge to $\mathbf{M}(a)$ as $n \to \infty$, $\Delta \to 0$ and $\varepsilon \to 0$.

It should be pointed out that the function $\sum_{i=1}^{n} \mathbf{1}_{X \in (a-\varepsilon, a+\varepsilon)}$ counts the number of observations inside the window $(a - \varepsilon, a + \varepsilon)$ and weights them equally. It seems plausible, however, that observations that are closer to a contain better information than more distant ones. In consequence, it might be worth replacing the so-called indicator kernel, i.e. $\mathbf{1}_{X \in (a-\varepsilon, a+\varepsilon)}$, with a function that is centered around a and converges to 0 as $|X| \to \varepsilon$. This is typically achieved by using smooth kernels $\mathbf{K}(.)$ that satisfy $\int \mathbf{K}(u)du = 1$ (c.f. Härdle and Linton (1994), for example, and the assumptions below). The ubiquitous normal kernel is an example. Hence, we can write

$$\widehat{\mathbf{M}}^{(2)}_{(n,\Delta,\varepsilon)}(a) = \frac{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right) \mathbf{f}(X_{i+\Delta}, X_i)/\Delta}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right)}, \tag{7}$$

which is a more general version of (6). As earlier, we expect $\widehat{\mathbf{M}}^{(2)}_{(n,\Delta,\varepsilon)}(a)$ to be consistent for $\mathbf{M}(a)$ as $n \to \infty$, $\Delta \to 0$ and $\varepsilon \to 0$.

We now summarize the features of the asymptotic requirements that appear to be necessary for consistency. In general, one needs to assume that the distance between observations Δ vanishes in the limit (i.e., infill asymptotics) while the time span (T, say) diverges to infinity (i.e., long span asymptotics) along with the number of observations n. As briefly mentioned in the Introduction and illustrated above in the context of simple examples, the former assumption (i.e., $\Delta \to 0$) is necessary to replicate the infinitesimal features of the theoretical quantities. The latter (i.e., $T, n \to \infty$) is crucial in order to guarantee that the number of visits that the sampled process makes in the neighborhood of a generic point a diverges to infinity in the limit (i.e., $\sum_{i=1}^{n} \mathbf{1}_{X_i \in (a-\varepsilon, a+\varepsilon)} \to \infty$ or $\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_i - a}{\varepsilon}\right) \to \infty$), provided the path of the underlying process does so. Of course, the additional assumption $\varepsilon \to 0$ permits proper conditioning at a asymptotically.

Coherently with our discussion, the following sampling scheme has been adopted by the recent literature on the fully functional estimation of continuous-time Markov processes and will be used throughout this Chapter. We will assume that we observe the process of interest X_t at points $\{t = t_1, t_2, .., t_n\}$ in the time interval $[0, T]$, with $T \geq T_0$, where T_0 and T are positive constants. Also, the data will be taken to be equispaced. Thus,

$$\{X_t = X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, X_{3\Delta_{n,T}}, ..., X_{n\Delta_{n,T}}\} \tag{8}$$

will be n observations at

$$\{t_1 = \Delta_{n,T}, t_2 = 2\Delta_{n,T}, t_3 = 3\Delta_{n,T}, ..., t_n = n\Delta_{n,T}\}, \tag{9}$$

where $\Delta_{n,T} = T/n$. In the limit, we will let $n \to \infty$, $T \to \infty$ and $\Delta_{n,T} = T/n \to 0$. In a few instances, T will be fixed at $\overline{T}$. In the sequel, we will be explicit about the limiting behaviour of the time span T.

Based on our discussion, it appears that the only requirements that we have to impose on the dynamic properties of the processes of interest for identification are those that guarantee divergence of the number of visits in the spatial vicinity of points in the range of the process. This is a typical feature of recurrent processes. Specifically, the sample path of a recurrent process returns to sets of non-zero Lebesgue measure an infinite number of times over time with probability one. We now rigorously state the definitions of recurrence that we will use in this review (the interested reader is referred to the standard treatment in Meyn and Tweedie (1993)).

Definition 1 (Null Harris Recurrence) *Let A be a measurable set of the range $\mathfrak{D}$ of the process of interest. Define the first hitting time of A as $\tau_A = \inf\{t \geq 0 : X_t \in A\}$. The process X_t is called null Harris recurrent if there is a $\sigma-$finite measure $m^*(dx)$ such that $m^*(A) > 0$ implies $\mathbf{P}^a[\tau_A < \infty] = 1$ for every $a \in \mathfrak{D}/\overline{A}$.*

Definition 2 (Positive Harris Recurrence) *Let A be a measurable set of the range $\mathfrak{D}$ of the process of interest. Define the first hitting time of A as $\tau_A = \inf\{t \geq 0 : X_t \in A\}$. The process X_t is called positive Harris recurrent (ergodic) if there is a $\sigma-$finite measure $m^*(dx)$ such that $m^*(A) > 0$ implies $\mathbf{E}^a[\tau_A] < \infty$ for every $a \in \mathfrak{D}/\overline{A}$.*

Define the occupation time measure of the set A of positive Lebesgue measure as

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds. \tag{10}$$

The quantity η_A^T gives the amount of time spent by the process at A between 0 and T. Under both notions of recurrence, we obtain $\mathbf{P}^a \left[\lim_{T\to\infty} \eta_A^T = \infty\right] = 1$ for $\forall a \in \mathfrak{D}/\overline{A}$. Specifically, starting from a level a not belonging to the generic set A, the process X_t returns to A an infinite number of times as $t \to \infty$ almost surely. This property is crucial for identification.

Null and positive recurrence are substantially milder assumptions than stationarity. Stationary processes are recurrent but recurrent processes do not have to be stationary. In particular, recurrent processes do not have to be endowed with a time-invariant stationary probability measure. Null recurrent processes, in fact, do not possess a time-invariant stationary probability measure. Nonetheless, null Harris recurrence implies the existence of a unique invariant measure $m(dx)$ $(= m^*(dx)$ in Definition 1) such that

$$m(A) = \int_{\mathfrak{D}} P\left(X_t^{(x)} \in A\right) m(dx) \quad \forall A \in \mathfrak{B}(\mathfrak{D}), \tag{11}$$

for every $0 \leq t < \infty$ (c.f. Azéma *et al.* (1967) and Karatzas and Shreve (1991), Exercise 6.18, page 362).[3] If the invariant measure is finite on $\mathfrak{D}$ ($m(\mathfrak{D}) < \infty$, that is), then the process is positive recurrent (ergodic) and has a time-invariant stationary probability measure (distribution, that is) to which it converges, at least in the limit. Such measure is given by $f(dx) = \frac{m(dx)}{m(\mathfrak{D})}$. A positive-recurrent process that is started in its stationary distribution $f(x)$ remains in the stationary distribution and, as a consequence, is stationary. Examples will be provided in the sequel. For now it suffices to say that Brownian motion in one (c.f. Example 1 in Section 3) and two dimensions are classical examples of null-recurrent processes. Conventional Vasicek (Ornestein-Uhlenbeck) and CIR processes (c.f. Vasicek (1977) and Cox *et al.* (1985)) are either positive recurrent (ergodic) or strictly stationary when initiated in their stationary measures (c.f. Example 5 in Section 3).

To conclude, the estimators that we review in this Chapter either follow the general form of (7) above or are constructed based on it. Null recurrence is all that we require to guarantee consistency of the estimates for the infinitesimal moments of interest. Positive recurrence and stationarity, which are clearly more stringent assumptions than null-recurrence, will be shown to determine only an increase in the rates of convergence of the estimators to the corresponding moments.

We now turn to a detailed analysis of the three processes that were mentioned in the introduction, namely scalar diffusion processes (SDP's), scalar jump-diffusion processes

[3] Alternatively, one could write

$$m = P_t m \qquad \forall t \geq 0,$$

where $(P_t)_{t\geq 0}$ is the semigroup of the process X_t (c.f. Ethier and Kurtz (1986) and Aït-Sahalia *et al.* (2001) in this volume).

(SJDP's) and multivariate diffusion processes (MDP's). A separate section will be devoted to each of these. In what follows we will not review the definitions of recurrence that were laid out previously but simply illustrate the conditions under which the processes of interest display stationary and/or recurrent behavior.

The following, rather standard, assumptions will be imposed on the kernel function $\mathbf{K}(.)$ throughout the present Chapter.

The kernel $\mathbf{K}(.)$ *is a continuously differentiable, symmetric and nonnegative function such that*

$$\int \mathbf{K}(s)ds = 1, \quad \int \mathbf{K}^2(s)ds < \infty \tag{12}$$

and

$$\int |s^2\mathbf{K}(s)|ds < \infty, \int |\mathbf{K}^{'}(s)|ds < \infty. \tag{13}$$

3 Scalar Diffusion Processes (SDP's)

In this section we model a generic time series as the solution X_t to the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t, \tag{14}$$

where B_t is a standard Brownian motion defined on the filtered probability space $(\Omega, \Im^B, (\Im_t^B)_{t\geq 0}, P)$. The initial condition $X_0 = \overline{X}$ belongs to L^2 and is taken to be independent of $\{B_t : t \geq 0\}$. Define the left-continuous filtration

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, B_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty, \tag{15}$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \textit{ with } N \subseteq G \textit{ and } P(G) = 0\}. \tag{16}$$

We create the augmented filtration

$$\widetilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \leq t < \infty, \tag{17}$$

and impose Assumption 1 through 3 (1 through 3bis) below, to assure the existence and pathwise uniqueness of a null recurrent (positive recurrent) and $\{\widetilde{\Im}_t^X\}$-adapted solution to (14).

(1) *$\mu(.)$ and $\sigma(.)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} = (l, u)$ with $-\infty \leq l < u \leq \infty$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions satisfy local Lipschitz and growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C_1|x - y|, \tag{18}$$

and

$$|\mu(x)| + |\sigma(x)| \leq C_2\{1 + |x|\}. \tag{19}$$

(2) *$\sigma^2(.) > 0$ on $\mathfrak{D}$.*

(3) *(Null recurrence) Define the second-order elliptic operator*[4]

$$\mathfrak{L}\varphi(.) = \varphi_x(.)\mu(.) + \frac{1}{2}\varphi_{xx}(.)\sigma^2(.). \tag{20}$$

There is a function $\varphi(.) : \mathfrak{R}\setminus\{0\} \to \mathfrak{R}$ of class C^2 in the domain of the operator that satisfies

$$\mathfrak{L}\varphi(.) \leq 0 \qquad on\ \mathfrak{R}\setminus\{0\} \tag{21}$$

and is such that $\Psi(r) := \min_{|x|=r} \varphi(.)$ is strictly increasing with $\lim_{r\to\infty} \Psi(r) = \infty$ (c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (i), page 370).

(3bis) *(Positive recurrence) There is a function $\varphi(.) : \mathfrak{R}\setminus\{0\} \to \mathfrak{R}$ of class C^2 in the domain of the operator that satisfies*

$$\mathfrak{L}\varphi(.) \leq -1 \qquad on\ \mathfrak{R}\setminus\{0\}, \tag{22}$$

and is such that $\Psi(r) := \min_{|x|=r} \varphi(.)$ is strictly increasing with $\lim_{r\to\infty} \Psi(r) = \infty$ (c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (iii), page 371).

Under Assumptions 1, 2 and 3 (3bis) the stochastic differential equation (14) displays a strong solution X_t that is unique, null recurrent (positive recurrent) and continuous in $t \in [0, T]$. In particular, the process X_t satisfies

[4] The operator $\mathfrak{L}$ is generally called the infinitesimal generator of the SDP X. We refer the interested reader to Aït-Sahalia *et al.* (2001), which is contained in the present volume, and to Hansen and Scheinkman (1995) for a discussion of estimation methods for stationary diffusions based on the properties of the infinitesimal generator.

$$X_t = \overline{X} + \int_0^t \mu(X_s)ds + \int_0^t \sigma(X_s)dB_s \tag{23}$$

almost surely, with $\int_0^t \mathbf{E}\left(X_s^2\right) ds < \infty$, and is a semimartingale. The dynamics of X_t are completely determined by the functions $\mu(.)$ and $\sigma(.)$. Such functions are typically the object of econometric interest.

Assumptions 3 and 3bis are vital in determining recurrent behavior for X_t. As pointed out earlier, null recurrence is a sufficient condition for the existence of σ-finite invariant measure $m(dx)$. Interestingly, such a measure is unique up to multiplication by a constant and is known to be equal (up to a proportionality factor) to the so-called *speed measure*, i.e.,

$$m(dx) = \frac{2dx}{S'(x)\sigma^2(x)} \qquad \forall x \in \mathfrak{D} \subseteq \mathfrak{R}, \tag{24}$$

where $S'(x)$ is the derivative of the *scale function*, namely

$$S(x) = \int_c^x \exp\left\{\int_c^y \left[-\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\} dy, \tag{25}$$

where $c \in \mathfrak{D}$. Under Assumption 3bis, the SDP is positive-recurrent (i.e., $m(\mathfrak{D}) < \infty$) and admits a time-invariant probability measure. In particular, the normalized speed measure, i.e. $m(dx)/m(\mathfrak{D})$, is the limiting stationary probability measure of X implying

$$\lim_{t\to\infty} P^x(X_t < z) = \frac{m((l,z))}{m(\mathfrak{D})} \qquad \forall x, z \in \mathfrak{D} \subseteq \mathfrak{R}, \tag{26}$$

(c.f. Pollack and Siegmund (1985) and Karatzas and Shreve, 1991, Exercise 5.40, page 353). More explictly, we can write the time-invariant stationary density of the process as

$$\begin{aligned} f(x) &= \frac{m(x)}{m(\mathfrak{D})} = \frac{1}{m(\mathfrak{D})}\frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)} \\ &= \left(\int_{\mathfrak{D}} \frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)} dx\right)^{-1} \frac{\exp\left\{\int_c^x \left[\frac{2\mu(s)}{\sigma^2(s)}\right] ds\right\}}{\sigma^2(x)}. \end{aligned} \tag{27}$$

We now provide some examples to fix the ideas.

Example 1 (Natural scale diffusions) *For general scalar diffusions, if the scale function $S(x)$ is such that $\lim_{x\to l+} S(x) = -\infty$ and $\lim_{x\to u-} S(x) = \infty$, then the process is recurrent, that is it satisfies $\mathfrak{L}\varphi(.) \leq 0$ where φ is defined in (3) and (3bis) above (c.f. Khasminskii (1980)).*

Apparently, the solution to $dX_t = \sigma(X_t)dB_t$ with σ continuous and strictly positive is Harris recurrent over $\Re$ with scale function $S(x) = x - c$ and invariant measure $m(dx) = \frac{2dx}{\sigma^2(x)}$. Chen, Hansen and Carrasco (1999), for example, discuss the mixing properties of the natural scale diffusion with $\sigma^2(x) = (1+x^2)^\gamma$ for $\frac{1}{2} < \gamma < 1$ (see Aït-Sahalia et al. (2001) in this volume). If $0 \leq \gamma \leq \frac{1}{2}$, then the process is null-recurrent on $\Re$. For values of γ strictly larger than $\frac{1}{2}$ the process is ergodic (positive recurrent). This is a case of "volatility-induced" reversion to the mean (c.f. Conley et al. (1998)). Trivially, standard Brownian motion (i.e., $\gamma = 0$) is null-recurrent.

Example 2 (Brownian motion with drift) *Assume X_t is the solution to $dX_t = \mu dt + \sigma dB_t$ with $\sigma > 0$. The scale function and the speed measure are $S(x) = \frac{1-e^{-\alpha(x-c)}}{\alpha}$ and $m(dx) = \frac{2e^{\alpha x}}{\sigma^2}dx$ where $\alpha = \frac{2\mu}{\sigma^2}$, respectively. If $\mu > 0$, then $\lim_{x\to\infty} S(x) = \frac{1-e^{\alpha c}}{2\alpha}$ and $\lim_{x\to-\infty} S(x) = -\infty$. The process is not recurrent and $P[\inf_{0\leq t<\infty} X_t > -\infty] = 1$. If $\mu < 0$, then $\lim_{x\to\infty} S(x) = \infty$ and $\lim_{x\to-\infty} S(x) = \frac{1-e^{\alpha c}}{2\alpha}$. The process is not recurrent and $P[\inf_{0\leq t<\infty} X_t < \infty] = 1$. In the former case X_t has an attracting boundary at ∞ ($P[\lim_{t\to\infty} X_t = \infty] = 1$, that is). In the latter case X_t has an attracting boundary at $-\infty$ ($P[\lim_{t\to\infty} X_t = -\infty] = 1$, that is). In both cases, it is easy to prove that the boundary is unattainable, i.e. it cannot be reached in finite time with positive probability (c.f. Karatzas and Shreve (1988) and Karlin and Taylor (1981)).*

Example 3 (Geometric Brownian motion) *Assume X_t is the solution to $dX_t = \mu X_t dt + \sigma X_t dB_t$ with $\mu, \sigma > 0$ and $\overline{X} > 0$. Then, $S(x) = c^\alpha \left[\frac{x^{-\alpha+1}}{-\alpha+1} - \frac{c^{-\alpha+1}}{-\alpha+1}\right]$ where $\alpha = \frac{2\mu}{\sigma^2}$ provided $\alpha < 1$ or $\alpha > 1$. The process is not recurrent for these choices of α. Specifically, if $\alpha < 1$, then $\lim_{x\to 0} S(x) = \frac{-c}{-\alpha+1}$ and $\lim_{x\to\infty} S(x) = \infty$ implying $P[\sup_{0\leq t<\infty} X_t < \infty] = 1 = P[\lim_{t\to\infty} X_t = 0]$. If $\alpha > 1$, then $\lim_{x\to\infty} S(x) = \frac{-c}{-\alpha+1}$ and $\lim_{x\to 0} S(x) = -\infty$ implying $P[\inf_{0\leq t<\infty} X_t > 0] = 1 = P[\lim_{t\to\infty} X_t = \infty]$. If $\alpha = 1$, then $S(x) = c[\log x - c]$ which implies $\lim_{x\to 0} S(x) = -\infty$ and $\lim_{x\to\infty} S(x) = \infty$, giving recurrence. In addition $m(dx) = \frac{2dx}{c\sigma^2 x}$ and is not integrable. Therefore, geometric Brownian motion is null-recurrent when $2\mu = \sigma^2$.*

Example 4 (Bessel process) *Assume $dX_t = \frac{d-1}{2X_s}dt + dW_t$ with $d \geq 2$ and $\overline{X} > 0$. If $d > 2$, we obtain $S(x) = c^{d-1}\left[\frac{c^{2-d} - x^{2-d}}{d-2}\right]$, giving $\lim_{x\to\infty} S(x) = \frac{c}{d-2}$ and $\lim_{x\to 0} S(x) = -\infty$. In consequence, the process is not recurrent and $P[\lim_{t\to\infty} X_t = \infty] = 1$. If $d = 2$, then $S(x) = c[\log x - \log c]$ implying recurrence. Furthermore, the speed measure (i.e., $m(dx) = \frac{2xdx}{c}$) is not integrable between 0 and $+\infty$ giving null-recurrence.*

Example 5 (Affine models) *Assume both the drift and the infinitesimal variance are linear functions of the state (i.e., $\mu(x) = c_0 + c_1 x$* and *$\sigma^2(x) = c_2 + c_3 x$ with $c_2, c_3 \geq 0$). The well-known Vasicek (Ornestein-Uhlenbeck) and CIR processes belong to this general* *class and are obtained by setting c_3 and c_2 equal to zero, respectively (c.f. Vasicek (1977) and Cox et al. (1985)). Under standard assumptions on the parameters (see Piazzesi (2001) in this volume for a thorough discussion of scalar and multivariate affine models and related estimation procedures), affine diffusions are strongly ergodic (positive recurrent). Should c_0 and c_1 be equal to zero and $\sigma^2(x) = c_2 + c_3|x|$ with $c_2 > 0$ and $c_3 \geq 0$, then the invariant measure would not be integrable over $\Re$ and the resulting process would be null recurrent (c.f. Example 1).*

Prior to describing the estimation strategy, we wish to discuss decriptive tools that have been recently introduced to characterize recurrent SDP's. Such tools rely on the notion of local time (Protter (1995) and Revuz and Yor (1998) are classical references). Local time is a random quantity that measures the amount of time that the process spends in the vicinity of a point. As a consequence, it might be intepreted as a spatial density and might be used to analyse the locational features of a possibly non-stationary SDP's in just the same way as a time-invariant stationary density can be used to study stationary processes (c.f. Phillips (1998, 2001)). The next subsection defines local time and introduces a simple estimation strategy to identify it based on a discrete sample of observations. We will also discuss the role that estimated local time can play as a descriptive statistic for recurrent SDP's and its importance in designing robust (to deviations from stationarity) identification procedures for processes whose dynamic is driven by (14) (c.f. Bandi (1998)). The terms 'local time', 'spatial density' and 'sojourn time' will be used interchangeably in the sequel.

3.1 Generalized Density Estimation for SDP's

The local time of a continuous semimartingale is defined as the random quantity $L_X(t, a)$ that satisfies

$$L_X(t, a) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbf{1}_{[a,\ a+\varepsilon[}(X_s) d[X]_s \ \forall a, t \tag{28}$$

where $[X]_t$ is the quadratic variation process of the underlying continuous semimartingale at t. Formula (28) clarifies the sense in which $L_X(t, a)$ measures time in information units or, more rigorously, in units of the quadratic variation process. Interestingly, one could write

$$[X]_t = \int_{-\infty}^{\infty} L_X(t, a) da, \tag{29}$$

thus expressing the quadratic variation process in terms of contributions coming from fluctuations in the process that occur in the vicinity of different spatial points $a \in (-\infty, \infty)$. Equation (29) is a spatial decomposition of variation. Readers familiar with stationary time series analysis will recognise the similarity of (29) to the decomposition of the variance σ_X^2 of a process X in terms of its spectral density at different frequencies, i.e.,

$$\sigma_X^2 = \int_{-\pi}^{\pi} f_{xx}(w)dw, \tag{30}$$

where $f_{xx}(w)$ is the spectral density of X.

We can specialize the analysis to the case of SDP's and consider a rescaled version of the standard notion of sojourn time defined as

$$\overline{L}_X(t,a) = \frac{L_X(t,a)}{\sigma^2(a)}. \tag{31}$$

Since $d[X]_s = \sigma^2(X_s)ds$ when assuming that the underlying semimartingale is an SDP as in (23), then formula (31) can be interpreted as representing time in real time units rather than in units of the non-decreasing quadratic variation process. In other words, $\overline{L}_X(t,a)$ records the amount of calendar time spent by the process in the neighborhood of a and can be defined as 'chronological local time' (c.f. Bosq (1998) and Park and Phillips (1998)). Such a notion has an interesting interpretation. Consider the occupation measure that was introduced in Section 2, namely

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}). \tag{32}$$

As pointed out earlier, the quantity η_A^T represents the amount of time spent by the process in a certain spatial set of nonzero Lebesgue measure. Chronological local time is nothing but the density of the occupation measure of the process. Put differently, chronological local time is a version of the Radon-Nikodym derivative of the occupation measure with respect to the Lebesgue measure $\left(\frac{\partial \eta^T}{\partial s}, \text{that is}\right)$ and is an occupation density (c.f. Geman and Horowitz (1980)). In fact, we can write

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds = \int_A \overline{L}_X(T,a)da, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}) \tag{33}$$

and, by linearity and monotone convergence,

$$\int_0^T \Phi(X_s)ds = \int \Phi(a)\overline{L}_X(T,a)da, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}) \tag{34}$$

where Φ is a Borel measurable, non-negative, function (c.f. Bosq (1998), *inter alia*). Formula (34) is typically called the 'occupation time formula' and can be regarded as the analogue of a standard expectation (i.e. the integral with respect to a time-invariant probability measure) in the analysis of times series that are not necessarily endowed with a time-invariant probability measure (c.f. Phillips and Park (1998) and Phillips (2001)).

From an applied standpoint, the notion of chronological local time is relevant for at least three, mutually reinforcing, reasons. First, chronological local time has an appealing intepretation in terms of calendar time spent by the process in the vicinity of values in its range. Second, chronological local time arises naturally as the limiting process to which density-like kernel estimators converge provided the underlying process is a scalar semimartingale and suitable conditions on the bandwidth used are met. Third, as shown in the next subsection, this is the notion of local time that will play a crucial role in understanding the limiting distributions of the infinitesimal moments $\mu(.)$ and $\sigma^2(.)$. In what follows, we will use the convention of referring to it simply as 'local time.'

The first two reasons together suggest the usefulness of local time as a new method for the descriptive analysis of data that might not be stationary, so that the techniques can be used in situations where estimated probability density functions do not make theoretical sense. Recent work has proposed nonparametric estimates of the local time process in this way and interpreted them in terms of generalized densities to be employed as new descriptive tools for studying the spatial characteristics of time series that may be nonstationary. The original intuition is due to Phillips (1998) in the context of nonstationary discrete-time series that can be embedded in Brownian motion. In continuous-time finance models, local time was first used as a descriptive tool for potentially nonstationary SDP's of the form analysed here by Bandi (1998).

As pointed out earlier, a natural way to identify local time is to use density-like kernel estimators. Based on the same sampling scheme as in Section 2 with $T = \overline{T}$, we define an estimate of $\overline{L}_X(\overline{T}, a)$ as

$$\widehat{\overline{L}}_X(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right), \tag{35}$$

where $h_{n,\overline{T}}$ is bandwidth sequence depending on n and $\mathbf{K}(.)$ is a conventional kernel function that satisfies the assumptions in Section 2. Theorems 3.1 and 3.2 below show consistency of the local time estimator for its theoretical counterpart and provide a limiting distribution.

Theorem 3.1 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(= \overline{T})$ in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, then*

$$\widehat{\overline{L}}_X(\overline{T}, a) \stackrel{a.s.}{\to} \overline{L}_X(\overline{T}, a) \qquad \forall a \in \mathfrak{D} \tag{36}$$

Proof *See Florens-Zmirou (1993). For a proof that allows for more general kernel functions than the indicator kernel (as in Florens-Zmirou (1993)) and utilizes different statistical tools, such as the occupation time formula in (34), see BP (1998).*

We now turn to the asymptotic distribution.

Theorem 3.2 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(=\overline{T})$ in such a way that $\frac{1}{h_{n,\overline{T}}^{3/2}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, then*

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{L}}_X(\overline{T},a) - \overline{L}_X(\overline{T},a)\right) \Rightarrow \mathbf{MN}\left(0, 8\mathbf{k}\frac{1}{\sigma^2(a)}\overline{L}_X(\overline{T},a)\right) \qquad \forall a \in \mathfrak{D} \tag{37}$$

where $\mathbf{k} = \int_0^\infty \int_0^\infty \min(s,q)\mathbf{K}(s)\mathbf{K}(q)dsdq$.

Proof *Phillips (1998) gives a very similar result for the estimated Brownian local time of linear and nonstationary discrete-time series that are embeddable in Brownian motion. The reader is referred to Bandi (1998) for a proof in the case of an underlying SDP that is assumed to be the unique and strong solution to a stochastic differential equation like (14), coherently with the statement of the theorem.*

Theorem 3.1 justifies estimating the calendar time that an SDP spends in the local vicinity of a point by using a density-like kernel estimator. Theorem 3.2 enables us to construct asymptotic confidence intervals which closely resemble conventional intervals for probability densities obtained from kernel estimates (c.f. Phillips (1998) and Bandi (1998)). In fact, the (estimated) asymptotic 95% confidence interval of $\overline{L}_X(\overline{T},a)$ is given by

$$\widehat{\overline{L}}_X(\overline{T},a) \pm 1.96\left(8\mathbf{k}\frac{h_{n,\overline{T}}}{\sigma^2(a)}\widehat{\overline{L}}_X(\overline{T},a)\right)^{1/2}. \tag{38}$$

It is worth recalling here that the limiting process $\overline{L}_X(\overline{T},a)$ is random. As opposed to standard probability densities, spatial densities have a time dimension that can be fruitfully explored by changing the span of data used in the implementation of (35). In other words, $\widehat{\overline{L}}_X(\overline{T}_1,a)$ and $\widehat{\overline{L}}_X(\overline{T}_2,a)$ measure the 'estimated' time spent by the SDP of interest at a in the time intervals $[0,\overline{T}_1]$ and $[0,\overline{T}_2]$, respectively, and can be used as robust (to deviations from stationarity) descriptive statistics to summarize the spatial evolution of the SDP over time.

A couple of additional observations are in order. Given the interpretation of local time, the following result should come as no surprise.

Theorem 3.3 *Assume X_t the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3bis)). If $h_{n,T} \to 0$ as $n, T \to \infty$ in such a way that $\frac{T}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$, then*

$$\frac{\widehat{\overline{L}}_X(T,a)}{T} \overset{a.s.}{\to} f(a) = \frac{m(a)}{m(\mathfrak{D})} \qquad \forall a \in \mathfrak{D} \tag{39}$$

Proof *See BP (1998) and Moloche (2000). The interested reader is also referred to Moloche (2000) for a discussion of the limiting properties of the expected local time process.*

This result simply tells us that the standardized local time estimator of a strictly stationary (or positive recurrent) SDP converges pointwise to the stationary density of the process with probability one. Very loosely speaking, if we divide the estimated time spent by the process at a between 0 and T by T, by appealing to the conventional frequentist notion of probability, we expect the ratio to converge to the probability mass at a when letting T diverge to infinity. Equivalently, we can say that the local time of a stationary, or positive recurrent, process diverges to infinity linearly with T (see, also, Bosq (1997) and Bosq and Davidov (1998)). Naturally, we expect nonstationary but recurrent processes to have local times that diverge at speeds that are slower than T. Such speeds are generally not quantifiable for general nonstationary, recurrent processes. Nonetheless, Brownian motion is known to have a local time that diverges at speed $\sqrt{T}$. The following result can be easily proved for a standard Brownian motion (c.f. BP (1998) and Moloche (2000) for a more general method of proof):

$$\frac{\widehat{\overline{L}}_W(T,a)}{\sqrt{T}} = \frac{\sqrt{T}L_W\left(1, \frac{a}{\sqrt{T}}\right)}{\sqrt{T}} + o_{a.s.}(1) \overset{a.s.}{\to} L_W(1,0). \tag{40}$$

We come back to a discussion of the divergence rates of local time when describing the estimation procedure for drift and diffusion function. For the time being, it suffices to stress that the class of SDP's that we are studying, namely the class of recurrent SDP's, has local times that diverge to infinity with probability one when the time span does so. The reason is easy to explain. Local time measures the time spent by the process at a point between 0 and T, say. Scalar recurrent processes visit every point an infinite number of times as T goes to infinity with probability one. Necessarily, therefore, the local time of a recurrent process diverges to infinity almost surely as T diverges to infinity.

We complete this subsection by pointing out that functions of spatial densities can be used as descriptive tools for possibly nonstationary SDP's just like functions of probability densities are employed as descriptive statistics in the context of stationary time-series (c.f. Phillips (1998)). For example, Phillips (1998) defines a new kind of hazard function for discrete-time nonstationary time series as

$$\overline{H}_X(T,a) = \frac{\overline{L}_X(T,a)}{\int_a^\infty \overline{L}_X(T,s)ds} \tag{41}$$

where, as usual, $\overline{L}_X(T,a)$ is the standard sojourn time. Such a function can be interpreted as the spatial counterpart of conventional hazard functions (c.f. Silverman (1986) and Prakasa-Rao (1983)) where probability densities replace local times and might be used to quantify hazards of certain financial time series such as inflation rates or interest rates.

When applied to interest rates (c.f. Bandi (1998)), for instance, formula (41) gives the conditional risk over the period $[0, T]$ of an interest rate level of a, given that interest rates are at least as big as a. An asymptotic theory for kernel estimates of spatial hazard rates is available to assist statistical inference.

Theorem 3.4 *Assume X_t is the solution to (14). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(= \overline{T})$ in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,T}))^{1/2} = o(1)$, then*

$$\widehat{\overline{H}}_X(\overline{T}, a) = \frac{\widehat{\overline{L}}_X(\overline{T}, a)}{\int_a^\infty \widehat{\overline{L}}_X(\overline{T}, s)ds} \overset{a.s.}{\to} \overline{H}_X(\overline{T}, a) \qquad \forall a \in \mathfrak{D}. \tag{42}$$

Furthermore, if $h_{n,\overline{T}} \to 0$ as $n \to \infty$ for a fixed time span T $(= \overline{T})$ and

$$\frac{1}{h_{n,\overline{T}}^{3/2}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,T}))^{1/2} = o(1),$$

then

$$\frac{1}{\sqrt{h_{n,\overline{T}}}}\left(\widehat{\overline{H}}_X(\overline{T}, a) - \overline{H}_X(\overline{T}, a)\right) \Rightarrow \mathbf{MN}\left(0, \frac{8\mathbf{k}(\overline{H}_X(\overline{T}, a))^2}{\sigma^2(a)\overline{L}_X(\overline{T}, a)}\right) \qquad \forall a \in \mathfrak{D}, \tag{43}$$

where $\mathbf{k} = \int_0^\infty \int_0^\infty \min(s, q)\mathbf{K}(s)\mathbf{K}(q)dsdq$.

Proof *As in the case of Theorem 3.3, Phillips (1998) derives a similar result for the estimated spatial hazard function of linear and nonstationary discrete-time series that are embeddable in Brownian motion. The reader is referred to Bandi (1998) for a proof in the case of an underlying SDP that is assumed to be the unique and strong solution to a stochastic differential equation like (14), consistently with the statement of the theorem.*

Estimated spatial densities and spatial hazard rates have been used by Bandi (1998) in studying the temporal dynamics of a scalar diffusion model for the short-term interest rate process coherent with (14) above. We refer the interested reader to that paper for a discussion of the empirical implementation of the methodology in the case of SDP's. For an introduction to descriptive methods for nonstationary time series, the reader is referred to Phillips (1998, 2001). We now turn to the estimation of the infinitesimal conditional moments $\mu(.)$ and $\sigma^2(.)$.

3.2 Kernel Estimation of the Infinitesimal Moments of an SDP

It is well-known that the transition density of the unique solution to (14) is completely characterized by the functions $\mu(.)$ and $\sigma^2(.)$. In other words, understanding the temporal evolution of a general SDP amounts to indentifying the drift and the diffusion function. As discussed in Section 2, such functions have (infinitesimal) conditional moment definitions, i.e.

$$\begin{aligned} \mathbf{E}^a[X_t - a] &= t\mu(a) + o(t) && (44)\\ \mathbf{E}^a[(X_t - a)^2] &= t\sigma^2(a) + o(t), && (45) \end{aligned}$$

as $t \downarrow 0$, or

$$\begin{aligned} \mu(a) &= \lim_{t\to 0} \frac{1}{t}\mathbf{E}^a[X_t - a], && (46)\\ \sigma^2(a) &= \lim_{t\to 0} \frac{1}{t}\mathbf{E}^a\left[(X_t - a)^2\right], && (47) \end{aligned}$$

where a is a generic initial condition and $\mathbf{E}^a$ is the expectation operator associated with the process started at a. Loosely speaking, (44) and (45) can be interpreted as representing the "instantaneous" conditional mean and the "instantaneous" conditional variance of the process when $X_0 = a$.

Our informal arguments in Section 2, combined with the definitions of $\mu(.)$ and $\sigma^2(.)$ from (44) and (45) above, suggest that standard functional techniques for conditional expectations based on local averages are natural tools to estimate the two functions that drive the evolution of a general SDP. This is the intuition in BP (1998) where sample analogues to the infinitesimal conditional expectations are used to estimate the theoretical drift and diffusion function. Consider the same sampling scheme as in Section 2 (i.e. $n, T \to \infty$ with $\frac{T}{n} \to 0$). Coherently with Section 2, define

$$\begin{aligned} \widehat{\mu}_{(n,T)}(a) &= \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)}{\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)}, && (48)\\ \widehat{\sigma}^2_{(n,T)}(a) &= \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^2}{\sum_{i=1}^n \mathbf{K}\left(\frac{X_{i\Delta_{n,T}}-a}{h_{n,T}}\right)}, && (49) \end{aligned}$$

where $\mathbf{K}(.)$ is a kernel function satisfying the assumptions in Section 2. Formulae (48) and (49) can be interpreted as the Nadaraya-Watson kernel estimates corresponding to (46) and (47) above and they belong to a general class of estimates suggested in BP (1998). We now consider some aspects of the asymptotic theory in that paper. We start with (48).

Theorem 3.5 *Assume X_t is the solution to (14). Also, assume $h_{n,T}$ is such that*

$$\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1)$$

and $h_{n,T}\overline{L}_X(T,a) \overset{a.s.}{\to} \infty$ $\forall a \in \mathfrak{D}$ *as* $n, T \to \infty$ *with* $\frac{T}{n} \to 0$. *Then,*

$$\widehat{\mu}_{(n,T)}(a) \overset{a.s.}{\to} \mu(a). \tag{50}$$

Furthermore, if $h^5_{n,T}\overline{L}_X(T,a) = O_{a.s.}(1)$ $\forall a \in \mathfrak{D}$, *then*

$$\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\left\{\widehat{\mu}_{(n,T)}(a) - \mu(a) - \Gamma_\mu(a)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\sigma^2(a)\right), \tag{51}$$

where

$$\Gamma_\mu(a) = h^2_{n,T}\mathbf{K}_1\left[\mu^{'}(a)\frac{m^{'}(a)}{m(a)} + \frac{1}{2}\mu^{''}(a)\right], \tag{52}$$

$m(a)$ *is the speed function of the process* X *at* a, *i.e.* $m(a) = \frac{2}{S'(a)\sigma^2(a)}$, $\mathbf{K}_1 = \int_{-\infty}^{\infty} s^2\mathbf{K}(s)ds$ *and* $\mathbf{K}_2 = \int_{-\infty}^{\infty}\mathbf{K}^2(s)ds$.

Proof *See BP (1998).*

We now turn to (49).

Theorem 3.6 *Assume* X_t *is the solution to (14). Also, assume* $h_{n,T}$ *is such that*

$$\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1) \qquad \forall a \in \mathfrak{D}$$

as $n, T \to \infty$ *with* $\frac{T}{n} \to 0$. *Then,*

$$\widehat{\sigma}^2_{(n,T)}(a) \overset{a.s.}{\to} \sigma^2(a). \tag{53}$$

Furthermore, if $\frac{h^5_{n,T}\overline{L}_X(T,a)}{\Delta_{n,T}} = O_{a.s.}(1)$ $\forall a \in \mathfrak{D}$, *then*

$$\sqrt{\frac{h_{n,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{n,T}}}\left\{\widehat{\sigma}^2_{(n,T)}(a) - \sigma^2(a) - \Gamma_{\sigma^2}(a)\right\} \Rightarrow \mathbf{N}\left(0, 4\mathbf{K}_2\sigma^4(a)\right), \tag{54}$$

where

$$\Gamma_{\sigma^2}(a) = h^2_{n,T}\mathbf{K}_1\left[\left(\sigma^2(a)\right)^{'}\frac{m^{'}(a)}{m(a)} + \frac{1}{2}\left(\sigma^2(a)\right)^{''}\right], \tag{55}$$

$m(a)$ *is the speed function of the process* X *at* a, *i.e.* $m(a) = \frac{2}{S'(a)\sigma^2(a)}$, $\mathbf{K}_1 = \int_{-\infty}^{\infty} s^2\mathbf{K}(s)ds$ *and* $\mathbf{K}_2 = \int_{-\infty}^{\infty}\mathbf{K}^2(s)ds$.

Proof *See BP (1998).*

Under appropriate conditions on the bandwidths, the estimators converge to the true functions with probability one. The asymptotic distributions are normal and centered on the relevant functions provided the bandwidth sequences converge to zero sufficiently fast. If this is not the case, then non-random bias terms affect the limiting distributions. The diffusion estimator converges to its theoretical counterpart at a faster rate $\left(\text{i.e. } \sqrt{\frac{h_{n,T}\widehat{\overline{L}}_X(T,a)}{\Delta_{n,T}}}\right)$ than the drift estimator $\left(\text{i.e. } \sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\right)$. We will now be more specific about the drift case but similar arguments apply to the diffusion function. A discussion of the difference between the two cases will follow.

When dealing with the drift, local time plays the same role that is played by the number of observations in the more standard estimation of conditional moments in discrete-time. What matters to identify the drift at a point a is not the rate of divergence of the number of data points n but the rate of divergence of the number of calendar time units spent by the process in the vicinity of the level a (c.f. Section 2). Not surprisingly, therefore, the standard condition $nh_n \rightarrow \infty$ is replaced in our case by $h_{n,T}\overline{L}_X(T,a) \stackrel{a.s.}{\rightarrow} \infty \; \forall a \in \mathfrak{D}$ as $n, T \rightarrow \infty$ with $\frac{T}{n} \rightarrow 0$. Equivalently, the pointwise condition that needs to be imposed on the bandwidth to prevent the insurgence of a bias term in the limit is $h_{n,T}^5\overline{L}_X(T,a) \stackrel{a.s.}{\rightarrow} 0$ as opposed to the more conventional assumption $h_n^5 n \rightarrow 0$. In other words, the smoothing parameter has to converge to zero slowly enough as to guarantee that $h_{n,T}\overline{L}_X(T,a) \stackrel{a.s.}{\rightarrow} \infty$ (rather than $nh_n \rightarrow \infty$) but sufficiently fast as to satisfy $h_{n,T}^5\overline{L}_X(T,a) \stackrel{a.s.}{\rightarrow} 0$ (rather than $h_n^5 n \rightarrow 0$).

Correspondingly, the rate of convergence of the estimator is random and equal to $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$ rather than $\sqrt{nh_n}$. Let us now consider the asymptotic variance and bias. These are given by

$$\frac{\left(\int_{-\infty}^{\infty} \mathbf{K}^2(s)ds\right)\sigma^2(a)}{h_{n,T}\widehat{\overline{L}}_X(T,a)} \tag{56}$$

and

$$h_{n,T}^2\left(\int_{-\infty}^{\infty} s^2\mathbf{K}(s)ds\right)\left[\mu^{'}(a)\frac{m^{'}(a)}{m(a)} + \frac{1}{2}\mu^{''}(a)\right], \tag{57}$$

respectively. Their interpretation is clear when considering well-known findings about the asymptotic bias and variance of the standard Nadaraya-Watson estimator of conditional moments in discrete time (see, for instance, formula (3.60) and theorem 3.5 in Pagan and Ullah (1999)). In particular, the spatial density $\overline{L}_X(T,a)$ and the ratio between the derivative of the speed function and the speed function itself play the same role as that played by the density $f(a)$ and the ratio between the derivative of the density function and the density function itself in conventional nonparametric analysis. The features of this theory are a reflection of the generality of the assumptions imposed on the underlying process. As pointed out earlier, recurrence is all that is required.

In consequence, the theory is specializable to the positive recurrent and stationary cases. The following theorems mirror more conventional results in the functional estimation of conditional expectations for stationary, discrete-time series, and are immediate after noticing that

$$\frac{\widehat{\overline{L}}_X(T,a)}{T} \overset{a.s.}{\to} f(a) \tag{58}$$

and[5]

$$\frac{m'(a)}{m(a)} = \frac{f'(a)}{f(a)} = (\log f(a))' = \frac{2\mu(a) - \left(\sigma^2(a)\right)'}{\sigma^2(a)} \tag{59}$$

under positive recurrence or stationarity (c.f. the discussion in the previous subsection).

Thereom 3.7 *Assume X_t is the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3bis)). Furthermore, assume $h_{n,T} \to 0$ as $n, T \to \infty$ with $\Delta_{n,T} \to 0$ so that $\frac{T}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$ and $h_{n,T}T \to \infty$. Then,*

$$\widehat{\mu}_{(n,T)}(a) \overset{a.s.}{\to} \mu(a) \qquad \forall a \in \mathfrak{D}. \tag{60}$$

Additionally,

$$\sqrt{h_{n,T}T}\left\{\widehat{\mu}_{(n,T)}(a) - \mu(a) - \Gamma_\mu(a)\right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\frac{\sigma^2(a)}{f(a)}\right), \tag{61}$$

if $h_{n,T} = O(T^{-1/5})$ where

$$\Gamma_\mu(a) = h_{n,T}^2\mathbf{K}_1\left[\mu'(a)\frac{f'(a)}{f(a)} + \frac{1}{2}\mu''(a)\right], \tag{62}$$

$f(a)$ is the stationary distribution function of the process at a, $\mathbf{K}_2 = \int_{-\infty}^{\infty}\mathbf{K}^2(s)ds$ and $\mathbf{K}_1 = \int_{-\infty}^{\infty} s^2\mathbf{K}(s)ds$.

Proof *See BP(1998).*

---

[5]The last equality in (59) can be proved by solving the equation

$$\int_{\mathfrak{D}} \mathfrak{L}\varphi(x)f(x)dx = 0$$

for a function $\varphi(.)$ in the domain of the infinitesimal generator $\mathfrak{L}$ (c.f. Aït-Sahalia *et al.* (2001) in this volume). Such an equation holds by stationarity (c.f. Hansen and Scheinkman (1995)).

Thereom 3.8 *Assume X_t is the solution to (14) and $m(\mathfrak{D}) < \infty$ (as implied by Assumption (3bis)). Furthermore, assume $h_{n,T} \to 0$ as $n, T \to \infty$ with $\Delta_{n,T} \to 0$ so that $\frac{T}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o(1)$. Then,*

$$\widehat{\sigma}^2_{(n,T)}(a) \overset{a.s.}{\to} \sigma^2(a) \qquad \forall a \in \mathfrak{D}. \tag{63}$$

Additionally,

$$\sqrt{nh_{n,T}}\left\{\widehat{\sigma}^2_{(n,T)}(a) - \sigma^2(a) - \Gamma_{\sigma^2}(a)\right\} \Rightarrow \mathbf{N}\left(0, 4\mathbf{K}_2\frac{\sigma^4(a)}{f(a)}\right), \tag{64}$$

if $h_{n,T} = O(n^{-1/5})$ where

$$\Gamma_{\sigma^2}(a) = h^2_{n,T}\mathbf{K}_1\left[\left(\sigma^2(a)\right)'\frac{f'(a)}{f(a)} + \frac{1}{2}\left(\sigma^2(a)\right)''\right], \tag{65}$$

$f(a)$ is the stationary distribution function of the process at a, $\mathbf{K}_2 = \int_{-\infty}^{\infty}\mathbf{K}^2(s)ds$ and $\mathbf{K}_1 = \int_{-\infty}^{\infty}s^2\mathbf{K}(s)ds$.

Proof *See BP (1998).*

The diffusion case deserves some additional observations. It was noted earlier that the rate of convergence of the drift estimator is $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$. This rate clarifies the sense in which identification of the infinitesimal first moment of an SDP requires an enlarging span of data (c.f. Geman (1979), *inter alia*). In fact, should T be fixed, then $\overline{L}_X(T,a)$ would be bounded in probability and the drift estimator would diverge at the rate $\sqrt{h_{n,T}}$ (c.f. Bandi (1998)). On the contrary, $\overline{L}_X(T,a)$ diverges to infinity as $T \to \infty$ by the assumption of recurrence (c.f. the previous subsection), thus ensuring that $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)} \overset{a.s.}{\to} \infty$ provided the bandwidth $h_{n,T}$ converges to zero slowly enough. The intuition why an enlarging span of data for drift estimation is needed was put forward in Section 2 but it is worth repeating here for clarity. In order to achieve consistency of the drift estimator at a certain spatial level, say a, we need the process to visit that level an infinite number of times over time. In this case we can take averages of the first differences between observations on the continuous path of the process that occur in the local neighborhood of a (c.f. Section 2) and hope to apply some law of large number. Recurrence guarantees that every level will be visited an infinite number of times over time provided $T \to \infty$. A diverging local time at a as $T \to \infty$ is simply the manifestation of the fact that, with probability one, the level a is crossed an infinite number of times, as the time span grows indefinitely.

Interestingly, the diffusion function can be identified over a fixed span of data (c.f. Florens-Zmirou (1993), Brugière (1991, 1993) and Jacod (1997), *inter alia*). In other words, we can use (49) above and fix $T = \overline{T}$ to derive limiting results.

Theorem 3.9 *Assume X_t is the solution to (14). Given $n \to \infty$, $T = \overline{T}$ and $h_{n,\overline{T}} \to 0$ as $n \to \infty$ so that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}}))^{1/2} = o(1)$, the estimator (49) converges to the true function with probability one.*

If $nh^4_{n,\overline{T}} \to 0$, then the asymptotic distribution of (49) is driven by a "martingale" effect and has the form

$$\sqrt{nh_{n,\overline{T}}}\left\{\widehat{\sigma}^2_{(n,\overline{T})}(a) - \sigma^2(a)\right\} \Rightarrow \mathbf{MN}\left(0, \frac{4\mathbf{K}_2\sigma^4(a)}{\overline{L}_X(\overline{T},a)/\overline{T}}\right), \tag{66}$$

where $\mathbf{K}_2 = \int_{-\infty}^{\infty} \mathbf{K}^2(s)ds$. If $nh^4_{n,\overline{T}} \to \infty$, then the asymptotic distribution of (49) is driven by a "bias" effect and has the form

$$\frac{1}{h^{3/2}_{n,\overline{T}}}\left\{\widehat{\sigma}^2_{(n,\overline{T})}(a) - \sigma^2(a)\right\} \Rightarrow \mathbf{MN}\left(0, 16\varphi^{ind}\frac{\left(\sigma^{'}(a)\right)^2}{\overline{L}_X(\overline{T},a)}\right), \tag{67}$$

where $\varphi^{ind} = 2\int_0^{\infty}\int_0^{\infty} \mathbf{K}(a)\mathbf{K}(b)\min(a,b)dadb$.

Proof *Florens-Zmirou (1993) was the first to use the estimator (49) above to identify the diffusion function. The kernel used in Florens-Zmirou (1993) is a discountinuous indicator kernel and the consistency proof is based on mean-squared deviation (see also Jacod (1997) for a refinement of the approach suggested by Florens-Zmirou (1993)). The asymptotic distribution (66) is discussed in Florens-Zmirou (1993) with $4\mathbf{K}_2 = 2$ due the nature of the kernel used. Jiang and Knight (1997) modify the Florens-Zmirou estimator and define it using a continuous kernel. Their consistency proof follows Florens-Zmirou (1992) and is in mean-squared. The weak convergence result in (66) is also discussed in Jiang and Knight (1997) where $4\mathbf{K}_2$ is set equal to 2 despite the fact that a general, potentially smooth, kernel is utilized. The statement of the theorem above is based on BP (1998) where almost sure convergence is proven. The treatment in BP (1998) hightlights the potential for a random bias term (i.e. (67) above) that might dominate the asymptotic distribution should the bandwidth sequence not converge to zero at a fast enough pace.*

We now turn to a brief discussion of bandwidth selection.

3.2.1 The Choice of the Bandwidth

Optimal bandwidth selection is technically very demanding in these models and represents an open field of research, no rigorous treatment being available at present, at least to the authors' knowledge. Based on Theorem 3.5, in the drift case one can write

$$h^{drift}_{n,T} = c^{drift}_{n,T}\frac{1}{\log \widehat{\overline{L}}_X(T,a)}\widehat{\overline{L}}_X(T,a)^{-1/5}, \tag{68}$$

where $\widehat{\overline{L}}_X(T,a)$ is the estimated local time at a and $c_{n,T}^{drift}$ is a constant of proportionality. Such an expression derives from the fact that the asymptotic mean-squared error (MSE) at a generic level a is of order

$$O_{a.s.}\left(\left(h_{n,T}^{drift}\right)^4\right) + O_{a.s.}\left(\frac{1}{h_{n,T}^{drift}\widehat{\overline{L}}_X(T,a)}\right) \tag{69}$$

and the best rate is obtained by taking $h_{n,T}^{drift} \propto \widehat{\overline{L}}_X(T,a)^{-1/5}$ in which case the limiting MSE is of order $\widehat{\overline{L}}_X(T,a)^{-4/5}$. Premultiplication by $\frac{1}{\log \widehat{\overline{L}}_X(T,a)}$ is useful to achieve a close-to-optimal rate of convergence, undersmoothing slightly and eliminating the influence of the non-random bias term from the asymptotic distribution of the drift estimates. As for the diffusion function, Theorem 3.6 and a similar argument to that above suggest the expression

$$h_{n,T}^{diff} = c_{n,T}^{diff}\frac{1}{\log\left(\widehat{\overline{L}}_X(T,a)/\Delta_{n,T}\right)}\left(\widehat{\overline{L}}_X(T,a)/\Delta_{n,T}\right)^{-1/5} \tag{70}$$

and, in consequence, the approximation

$$h_{n,T}^{diff} \approx c_{n,T}^{diff}\frac{1}{\log n}n^{-1/5}, \tag{71}$$

for a T that diverges to infinity sufficiently slowly. When T is fixed, as in Theorem 3.9, the previous condition becomes

$$h_{n,\overline{T}}^{diff} = c_{n,\overline{T}}^{diff}\frac{1}{\log n}n^{-1/4}. \tag{72}$$

Both (70) and (72) imply close-to-optimal rates, namely rates that almost maximize the speed of convergence of the proposed estimators to the functions of interest while preventing the insurgence of a bias term in the limit.

Some observations are in order. First, formula (68) suggests that there is explicit scope for local adaptation of the drift bandwidth sequence to the number of visits to the point at which estimation is performed. In fact, it appears that the optimal bandwidth for the drift should be smaller at levels that are often visited.[6] In light of the approximation (71) and the result in (72) such effect is more pronounced in the drift case than in the diffusion case. Second, Theorems 3.5 and 3.6 suggest that the optimal drift bandwidth is generally larger than the optimal diffusion bandwidth. The last two observations are reflections of the fact that the local dynamics of an SDP contain more information about the diffusion function than about the drift, thus rendering estimation of the infinitesimal second moment possible over a fixed span of data (see Bandi and Nguyen (1999) for a simulation study).

[6] Bias reduction is the standard justification for suggesting variable bandwidths whose magnitude is inversely related to the estimated density function (c.f. page 31 of Pagan and Ullah (1999)). Such justification is likely to be valid in our framework, thus reinforcing our previous conclusions.

Unfortunately, despite being widely used in empirical work, standard automatic technologies (c.f. Pagan and Ullah (1999)) to select the constants $c_{n,T}^{drift}$, $c_{n,T}^{diff}$ and $c_{n,\overline{T}}^{diff}$, such as least squares cross-validation, have not been justified in the case of SDP's. Nonetheless, contrary to drift estimation, conventional cross-validation procedures appear to perform reasonably well when dealing with diffusion function estimation (c.f. Bandi and Nguyen (1999)). Future work can usefully focus on the development of convincing criteria for determining the constants $c_{n,T}^{drift}$, $c_{n,T}^{diff}$ and $c_{n,\overline{T}}^{diff}$ in (68), (70), (71) and (72) above and the preliminary smoothing sequence that is required to define $\widehat{\widehat{L}}_X(T,a)$.

3.2.2 Extension in Kernel Estimation for SDP's

The estimators (48) and (49) belong to the general class of estimators suggested by BP (1998). Consistently with the discussion in BP (1998), one could envisage a more involved two-step procedure with further smoothing. First, we could define sample analogs to the values that drift and diffusion take at the sampled points, i.e.

$$\widetilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}})}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)}, \tag{73}$$

$$\widetilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}}) = \frac{1}{\Delta_{n,T}} \frac{\sum_{j=1}^{n-1} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)(X_{(j+1)\Delta_{n,T}} - X_{j\Delta_{n,T}})^2}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{X_{j\Delta_{n,T}}-X_{i\Delta_{n,T}}}{h_{n,T}}\right)}. \tag{74}$$

Second, estimated drift and diffusion values at the sampled points could be averaged using weights based on smooth kernels to recover the theoretical functions at levels that the sampled process does not visit, i.e.

$$\overline{\mu}_{(n,T)}(a) = \frac{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)\widetilde{\mu}_{(n,T)}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)}, \tag{75}$$

$$\overline{\sigma}^2_{(n,T)}(a) = \frac{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)\widetilde{\sigma}^2_{(n,T)}(X_{i\Delta_{n,T}})}{\sum_{i=1}^{n} \overline{\mathbf{K}}\left(\frac{X_{i\Delta_{n,T}}-a}{\varepsilon_{n,T}}\right)}. \tag{76}$$

with $\overline{\mathbf{K}}(.)$ possibly different from $\mathbf{K}(.)$ and $h_{n,T}$ possibly different from $\varepsilon_{n,T}$. Both $\overline{\mathbf{K}}(.)$ and $\mathbf{K}(.)$ satisfy the assumptions in Section 2.

Asymptotically, the doubly-smoothed estimates (75) and (76) offer additional flexibility over the simple counterparts in (48) and (49) above. In effect, they can improve the asymptotic trade-off between bias and variance effects, thus delivering smaller limiting MSE's than simple smoothing (c.f. BP (1998)). As usual, let us focus on the drift while keeping in mind that the intuition extends to the diffusion function.

If $h_{n,T}$ satisfies the conditions in Theorem 3.5 and $\varepsilon_{n,T}/h_{n,T} \to \phi$, then the limiting bias and variance of the drift estimator (75) at a generic point a are, from BP (1998),

$$\frac{\theta_\phi \sigma^2(a)}{h_{n,T}\widehat{\overline{L}}_X(T,a)}, \tag{77}$$

with

$$\theta_\phi = \int_{-\infty}^{\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \overline{\mathbf{K}}(a)\overline{\mathbf{K}}(e)\mathbf{K}(z-\phi e)\mathbf{K}(z-\phi a)dzdeda, \tag{78}$$

and

$$h_{n,T}^2 \mathbf{K}^\phi \left[\mu^{'}(a)\frac{m^{'}(a)}{m(a)} + \frac{1}{2}\mu^{''}(a)\right], \tag{79}$$

with

$$\mathbf{K}^\phi = \int_{-\infty}^{\infty} s^2 \mathbf{K}(s)ds + \phi \int_{-\infty}^{\infty} s^2 \overline{\mathbf{K}}(s)ds. \tag{80}$$

Clearly, if $\varepsilon_{n,T}/h_{n,T} \to \phi = 0$, then double-smoothing coincides asymptotically with single-smoothing by a straigtforward comparison of (77) and (79) with (56) and (57) above, respectively. Nonetheless, since θ_ϕ is a decreasing function of ϕ and $\mathbf{K}^\phi$ an increasing function of it, there is scope for using convoluted kernels in order to achieve asymptotic MSE's that are minimized at values ϕ that are strictly larger than zero.

In finite samples, the extra level of smoothing that is implied by the use of convoluted kernels might also improve the flexibility and performance of the estimation strategy. This is particularly true for the drift. The intuition is as follows. We stressed earlier that the optimal smoothing parameter for the drift is generally larger than the corresponding parameter for diffusion estimation. Nonetheless, there appears to be a fundamental difficulty in choosing the optimal drift bandwidth based on the estimated local time and, consequently, on the recurrence properties of the underlying process. The use of convoluted kernel functions can achieve in finite samples the level of smoothing for the drift that weighted averages based on simple kernels would guarantee with relatively more appropriate choices of the bandwidth. The cost that double-smoothing imposes in terms of oversmoothed infinitesimal second moments appears to be largely outweighted by its benefit for drift estimation (c.f. Bandi and Nguyen (1999)).

3.3 Local Linear Estimation of the Infinitesimal Moment of an SDP

It is immediate to see that the estimators (48) and (49) can be written as

$$\widehat{\mu}_{n,T}(a) = \arg\min_{\theta^\mu} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T}) - \theta^\mu\right\}^2 \tag{81}$$

and

$$\widehat{\sigma}^2_{n,T}(a) = \arg\min_{\theta^{\sigma^2}} \sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T})^2 - \theta^{\sigma^2}\right\}^2, \qquad (82)$$

respectively. Specifically, the Nadaraya-Watson estimates of the drift and diffusion function fit a constant line to the data in vicinity of the level a. Alternatively, one might think of fitting a polynomial locally and minimizing the criteria

$$\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T}) - \sum_{s=0}^{r}\theta^{\mu}_{r}(X_{i\Delta_{n,T}} - a)^r\right\}^2 \qquad (83)$$

and

$$\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left\{\frac{1}{\Delta_{n,T}}(X_{(i+1)\Delta_{n,T}} - X_{i\Delta n,T})^2 - \sum_{s=0}^{r}\theta^{\sigma^2}_{r}(X_{i\Delta_{n,T}} - a)^r\right\}^2 \qquad (84)$$

with respect to $\boldsymbol{\theta}^{\mu,\sigma^2}(a) = (\theta^{\mu,\sigma^2}_0, \theta^{\mu,\sigma^2}_1, ..., \theta^{\mu,\sigma^2}_r)(a)$. A simple argument based on Taylor expansions around a suggests that the proper estimates of $\mu(a)$ and $\sigma^2(a)$ are the first components $\widehat{\theta}^{\mu}_0(a)$ and $\widehat{\theta}^{\sigma^2}_0(a)$ of the estimated vectors $\widehat{\boldsymbol{\theta}}^{\mu}(a)$ and $\widehat{\boldsymbol{\theta}}^{\sigma^2}(a)$. The remaining minimizers $(\widehat{\theta}^{\mu,\sigma^2}_1, ..., \widehat{\theta}^{\mu,\sigma^2}_r)$ are estimates of the derivatives of the functions of interest (provided such derivatives exist). This approach was suggested by Moloche (2000), following standard work by Fan (1992, 1993) and Fan and Gijbels (1996) in nonparametric regression analysis (see also Pagan and Ullah (1999, p. 93)). Interestingly, the estimated vectors $\widehat{\boldsymbol{\theta}}^{\mu}(a)$ and $\widehat{\boldsymbol{\theta}}^{\sigma^2}(a)$ can be expressed in the form of regression estimates since the criteria (83) and (84) can be readily interpreted in terms of classical weighted least-squares problems. We only consider the drift case but a similar derivation applies to the diffusion function. Write

$$\mathbf{X}_{n,T}(a) = \begin{bmatrix} 1 & (X_{\Delta_{n,T}} - a) & ... & (X_1 - a)^r \\ & & ... & \\ 1 & (X_{(n-1)\Delta_{n,T}} - a) & ... & (X_{(n-1)\Delta_{n,T}} - a)^r \end{bmatrix}, \qquad (85)$$

$$\mathbf{y}_{n,T} = \begin{bmatrix} \frac{1}{\Delta_{n,T}}(X_{2\Delta_{n,T}} - X_1)^2 \\ ... \\ \frac{1}{\Delta_{n,T}}(X_{n\Delta_{n,T}} - X_{(n-1)\Delta_{n,T}})^2 \end{bmatrix}, \qquad (86)$$

and

$$\mathbf{W}_{n,T}(a) = diag\left(\frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{X_{1\Delta_{n,T}} - a}{h_{n,T}}\right), ..., \frac{\Delta_{n,T}}{h_{n,T}}\mathbf{K}\left(\frac{X_{(n-1)\Delta_{n,T}} - a}{h_{n,T}}\right)\right). \qquad (87)$$

The pointwise drift estimator $\widehat{\theta}_0^{\mu}(a)$ is the first component of the p-vector

$$\widehat{\boldsymbol{\theta}}^{\mu}(a) = (\mathbf{X}'_{n,T}(a)\mathbf{W}_{n,T}(a)\mathbf{X}_{n,T}(a))^{-1}\mathbf{X}'_{n,T}(a)\mathbf{W}_{n,T}(a)\mathbf{y}_{n,T}. \tag{88}$$

Moloche (2000) shows that under the same conditions on the bandwidth as in Theorem 3.5 and provided that the same sampling scheme as in Section 2 is adopted, the drift estimate $\widehat{\theta}_0^{\mu}(a)$ converges to to the true function with probability one and is normally distributed in the limit. Its asymptotic variance and bias have the form

$$\frac{\left(\int_{-\infty}^{\infty}\mathbf{K}^2(s)ds\right)\sigma^2(a)}{h_{n,T}\widehat{\overline{L}}_X(T,a)}, \tag{89}$$

and

$$h_{n,T}^2\left(\int_{-\infty}^{\infty}s^2\mathbf{K}(s)ds\right)\frac{1}{2}\mu''(a). \tag{90}$$

Formulae (89) and (90) should be compared to the corresponding quantities for the Nadaraya-Watson kernel estimate discussed earlier, namely (56) and (57) above. The comparison is standard and we refer the interested reader to the original work of Fan (1992) and the recent review of Pagan and Ullah (pages 104-106) for details. Here we simply stress that the variances are the same, but the biases are different. In particular, the bias of the local polynomial estimator does not depend on the ratio $\frac{m'(.)}{m(.)}$ and on the first derivative of μ at a and is 'design adaptive' in the sense of Fan (1992). A similar discussion applies to diffusion function estimator $\widehat{\theta}_0^{\sigma^2}(a)$ (Moloche (2000)).

3.4 Using Nonparametric Information to Estimate and Test Parametric Models for SDP's

It is natural to use the information contained in the nonparametric estimates to design more accurate parametric models and test parametric assumptions. BP (1999) discuss a simple semiparametric procedure to estimate potentially nonstationary diffusions that overcomes the usual inference problems posed by the unavailability of a closed-form expression for the transition density of the underlying process and does not require simulations. They consider a parametric class $(\boldsymbol{\theta}^{\mu}, \boldsymbol{\theta}^{\sigma}) = \boldsymbol{\theta} \in \Theta$ for the underlying SDP and compute the parameters of interest as

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{n,T}^{\mu} \quad : \quad &= \underset{\boldsymbol{\theta}^{\mu}\in\Theta^{\mu}\subset\Theta}{\arg\min}\; Q_{n,T}^{\mu} \\ &= \underset{\boldsymbol{\theta}^{\mu}\in\Theta^{\mu}\subset\Theta}{\arg\min}\;\frac{\overline{T}}{n}\sum_{i=1}^{n}\left(\widehat{\mu}_{(n,T)}\left(X_{i\Delta n,\overline{T}}\right) - \mu\left(X_{i\Delta n,\overline{T}}, \boldsymbol{\theta}^{\mu}\right)\right)^2,\end{aligned} \tag{91}$$

and

$$\widehat{\boldsymbol{\theta}}^{\sigma^2}_{n,T} \quad : \quad = \underset{\boldsymbol{\theta}^{\sigma^2}\in\Theta^{\sigma^2}\subset\Theta}{\arg\min} \; Q^{\sigma^2}_{n,T}$$
$$= \underset{\boldsymbol{\theta}^{\sigma^2}\in\Theta^{\sigma^2}\subset\Theta}{\arg\min} \; \frac{\overline{T}}{n}\sum_{i=1}^{n}\left(\widehat{\sigma}^2_{(n,T)}\left(X_{i\Delta n,\overline{T}}\right)-\sigma^2\left(X_{i\Delta n,\overline{T}},\boldsymbol{\theta}^{\sigma^2}\right)\right)^2, \qquad (92)$$

where $\widehat{\mu}_{(n,T)}\left(X_{i\Delta n,\overline{T}}\right)$ and $\widehat{\sigma}^2_{(n,T)}\left(X_{i\Delta n,\overline{T}}\right)$ are functional estimates (defined over an enlarging span of data for consistency – see Subsection 3.2) of the Nadaraya-Watson type (c.f. (48) and (49) above) at the i-th observation. The parameter values are chosen so that the averaged squared distance between the nonparametric curves at the sampled points and the adopted parametric specification is minimized. The asymptotic distributions of the parameter estimates are (variance mixtures of) normals and can be readily interpreted on the basis of well-known results for conventional non-linear least-squares problems. Nonetheless, the integrals that appear in the limiting variances are not integrals with respect to probability measures (i.e. expectations) but integrals with respect to local times (i.e., occupation integrals, c.f. (34)) due to the generality of the approach in the present context. By virtue of the averaging, the rates of convergence of the parameter estimates are faster than the rates of convergence of the functional estimators used to define (91) and (92) above (c.f. BP (1999)). This is a typical result in semiparametric problems (c.f., Andrews (1989), for example).

Apparently, the above criteria can be employed to test alternative parametric assumptions about the functions of interest. Consider the drift case. Assume one wishes to test the hypotheses $H_0 : \mu_0(x) = \mu(x,\boldsymbol{\theta}^{\mu})$ against $H_1 : \mu_0(x) \neq \mu(x,\boldsymbol{\theta}^{\mu})$. Provided a consistent (under the null) parametric estimate of $\boldsymbol{\theta}^{\mu}$, $\widetilde{\boldsymbol{\theta}}^{\mu}_{n,T}$ say, is obtained (the value $\widehat{\boldsymbol{\theta}}^{\mu}_{n,T}$ that minimizes (91) is a viable option) and the distribution of $\widehat{Q}^{\mu}_{n,T}(\widetilde{\boldsymbol{\theta}}^{\mu}_{n,T})$ is derived under the null, standard methods can be utilized to construct a consistent test. The use of nonparametric information to test parametric models based on the minimization of averaged squared errors like (91) and (92) above has a long history in hypothesis testing about density functions. Important early references in discrete-time are Bickel and Rosenblatt (1973), Rosenblatt (1975) and Hall (1984). More recently, Aït-Sahalia (1996) has applied the idea to the study of a stationary scalar diffusion model (consistent with (14) above) for the short-term interest rate process.

With the exception of the above mentioned paper, little work exists on parametric inference for generally specified null-recurrent diffusions. Recent research by Höpfner and Kutoyants (2001) discusses a method of inference for the parameter θ in the drift of the equation

$$dX_t = \theta\frac{X_t}{1+X_t^2}dt + \sigma dB_t,$$

where $\theta \in \Theta = \left(-\frac{\sigma^2}{2},\frac{\sigma^2}{2}\right)$. As they show, θ is the parameter which determines the speed of divergence of additive integrable functionals of the process (in the sense discussed in Section 5 - Theorem 5.1) and Θ is the maximal open interval over which the process is null-recurrent. Given knowledge of the diffusion function σ and availability of a continuum of observations, parametric estimation is conducted by maximum likelihood through conventional arguments based on change of measure.

4 Scalar Jump-Diffusion Processes (SJDP's)[7]

Throughout this section we model a time-series as the solution X_t of the stochastic differential equation with jumps

$$\begin{aligned} dX_t &= \left[\mu(X_{t-}) - \lambda(X_{t-})\int_Y g(X_{t-},y)\Gamma(dy)\right]dt + \sigma(X_{t-})dB_t + dJ_t \\ &= \left[\mu(X_{t-}) - \lambda(X_{t-})\mathbf{E}_Y[g(X_{t-},y)]\right]dt + \sigma(X_{t-})dB_t + dJ_t, \end{aligned} \tag{93}$$

where $\{B_t : t \geq 0\}$ and $\{J_t : t \geq 0\}$ are respectively a standard scalar Brownian motion and an independent jump process. The initial condition $X_0 = \overline{X}$ belongs to L^2 and is taken to be independent of both B_t and J_t.

The functions $\mu(.)$ and $\sigma(.)$ have the same interpretation as in scalar diffusion models (c.f. Section 3). Nonetheless, due to the presence of a discountinuous jump-component $dJ_t = \Delta X_t$, the path of X_t fails to be continuous in the state space as in the case of standard SDP's despite being right-continuous with left limits (*càdlàg*, that is). A brief description of the features of the jump component is in order.

We assume that the jumps are bounded (i.e. $\sup_t |\Delta X_t| \leq C < \infty$ almost surely where C is a non-random constant) and occur with conditional intensity $\lambda(.)$ (i.e. $\lambda(a)dt$ is the infinitesimal probability of a jump at the level a). The impact of a jump is given by the function $g(.,y)$ whose arguments are the level of the process and a generic random variable y which we assume to be endowed with the stationary measure $\Gamma(.)$. Specifically,

$$dJ_t = \Delta X_t = X_t - X_{t-} = \int_Y g(X_{t-},y)N(dt,dy), \tag{94}$$

where

$$N_t^{\Phi} = \sum_{j=1} \mathbf{1}_{[\tau_j \leq t, y_{\tau_j} \in \Phi]} \tag{95}$$

is a Poisson counting measure with stationary and independent increments (c.f. Protter (1995)). Conditions 1 through 4 (1 through 4bis) below are imposed on the model (c.f. BN (2000)).

[7] This section is a largely abridged and simplified version of BN (2000). We refer the interested reader to that paper for additional discussions, technical details and proofs.

(1) *The functions* $\mu(.)$, $\sigma(.)$, $g(.,.)$ *and* $\lambda(.)$ *are time-homogeneous and* $\mathfrak{B}$*-measurable on* $\mathfrak{D} = (l, u)$ *with* $-\infty \leq l < u \leq \infty$ *where* $\mathfrak{B}$ *is the* σ*-field generated by Borel sets on* $\mathfrak{D}$*. They are at least twice continuously differentiable. Hence, they satisfy local Lipschitz and growth conditions. Thus, for every compact subset* Ψ *of the domain of the process, there exist constants* C_1 *and* C_2 *such that, for all* x *and* z *in* Ψ*,*

$$|\mu(x) - \mu(z)| + |\sigma(x) - \sigma(z)| + \lambda(x) \int_Y |g(x, y) - g(z, y)|\Gamma(dy) \leq C_1|x - z|, \quad (96)$$

and

$$|\mu(x)| + |\sigma(x)| + \lambda(x) \int_Y |g(x, y)|\Gamma(dy) \leq C_2\{1 + |x|\}. \quad (97)$$

(2) *For a given* $\alpha > 2$*, there exists a constant* C_3 *such that*

$$\lambda(x) \int_Y |g(x, y)|^\alpha \Gamma(dy) \leq C_3\{1 + |x|^\alpha\}. \quad (98)$$

(3) $\lambda(.) > 0$ *and* $\sigma^2(.) > 0$ *on* $\mathfrak{D}$*.*

(4) *(Null-recurrence) Define the second order elliptic operator* $\mathfrak{L}$ *and the integro-differential operator* $\mathfrak{A}$ *of the continuous and discontinuous portions of the solution to (93) above as*

$$\mathfrak{L}\varphi(.) = \varphi_x(.)\mu(.) + \frac{1}{2}\varphi_{xx}(.)\sigma^2(.) \quad (99)$$

and

$$\mathfrak{A}\varphi(.) = \lambda(.) \int_Y [\varphi(. + g(., y)) - \varphi(.) - \varphi_x(.)g(., y)]\, \Gamma(dy), \quad (100)$$

respectively. Assume Φ *is a Borel measurable and bounded function on the closure* $\overline{A}$*. The exterior Dirichlet problem, i.e.*

$$(\mathfrak{L} + \mathfrak{A})u = 0 \qquad a.e. \qquad in\ \mathfrak{D} \setminus \overline{A} \quad (101)$$

$$u = \Phi \qquad a.e. \qquad in\ \overline{A} \quad (102)$$

admits a unique bounded solution $u(x)$ *(Menaldi and Robin (1999)).*

(4bis) *(Positive-recurrence) The exterior Dirichlet problem, i.e.*

$$-(\mathfrak{L} + \mathfrak{A})u = f \qquad a.e. \qquad in\ \mathfrak{D} \setminus \overline{A} \quad (103)$$

$$u = 0 \qquad a.e. \qquad in\ \overline{A} \quad (104)$$

admits a unique bounded solution $u(x)$ *(Menaldi and Robin (1999)).*

Under Assumptions 1 through 4 (4bis) the SJDP (93) has a strong solution that is unique and null-recurrent (positive-recurrent). In particular, the càdlàg process X_t satisfies

$$X_t = \overline{X} + \int_0^t \mu(X_{s-})ds + \int_0^t \sigma(X_{s-})dB_s + \int_{0+}^t \int_Y g(X_{s-}, y)\overline{\nu}(ds, dy) \tag{105}$$

where

$$\overline{\nu}(ds, dy) = N(ds, dy) - \nu(X_{s-}, dy)ds \tag{106}$$

$$= N(ds, dy) - \lambda(X_{s-})\Gamma(dy)ds \tag{107}$$

is a 'compensated' random measure and the notation $\int_{0+}^t = \int_{(0,t]}$ denotes the integral over the half open interval. It is noted that

$$\int_{0+}^t \int_Y g(X_{s-}, y)\overline{\nu}(ds, dy) \tag{108}$$

represents the conditional variation between 0+ and t in the path of the process due to discontinuous jumps of random size y (and impact $g(., y)$) net of its expected conditional magnitude at $0+$. The model is defined as 'compensated' by virtue of the presence of the term $\lambda(X_t)\mathbf{E}_Y[g(X_t, y)]$ denoting the conditional mean of the jump part. Its presence ensures that this component has the martingale property and the solution to the SJDP (93) is a semimartingale. The semimartingale property, which is trivially satisfied by standard SDP's, makes SJDP's of the kind analyzed here attractive for modelling purposes in continuous-time finance. In fact, this property implies the existence of an equivalent martingale measure under which the process is a (local) martingale and, more importantly, absence of arbitrage in the spaces that preclude *doubling strategies* (c.f. Duffie (1996)).

Given Assumptions (1), (2) and (3), the infinitesimal conditional moments of the changes in the solution to (93) above can be written in terms of the functions $\mu(.)$, $\sigma(.)$, $g(.,.)$ and $\lambda(.)$ (c.f. Gikhman and Skorohod (1972)), i.e.

$$\mathbf{M}^1(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}^a[X_t - a] = \mu(a), \tag{109}$$

$$\mathbf{M}^2(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}[(X_t - a)^2] = \sigma^2(a) + \lambda(a)\mathbf{E}_Y[g^2(a, y)], \tag{110}$$

$$\mathbf{M}^k(a) = \lim_{t \to 0} \frac{1}{t}\mathbf{E}[(X_t - a)^k] = \lambda(a)\mathbf{E}_Y[g^k(a, y)] \ \forall k > 2, \tag{111}$$

for a generic $a \in \mathfrak{D}$. Specifically, formulae (109) through (111) clarify the sense in which the conditional moments of an SJDP contain information about the intensity of a jump and the distribution of the jump component as well as about the drift $\mu(.)$ and the so-called diffusive volatility $\sigma^2(.)$. This observation led Johannes (2000) to suggest nonparametric estimates of the infinitesimal moments and a procedure to extract the parameters and functions of interest from the estimated moments. We will be more accurate in the sequel. For now it suffices to point out that the $\mathbf{M}^k(.)$'s will be our object of econometric interest (c.f. BN (2000)).

We now turn to generalized density estimation for processes that are possibly nonstationary solutions to stochastic differential equations with jumps such as (93) above.

4.1 Generalized Density Estimation for SJDP's

The theory of local times for càdlàg semimartingales is well established in the stochastic process literature. We refer the reader to Protter (1995) for a thorough treatment. Here, consonant with our discussion in Section 3, we review some basic notions that will serve the purpose of illustrating the role that estimated local time can play as a descriptive tool for recurrent SJDP's. Assume X_t is a càdlàg semimartingale, then its sojourn time at T and a can be written as

$$L_X(T,a) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^T \mathbf{1}_{[a,\ a+\varepsilon[}(X_s) d[X]_s^c \qquad a.s., \tag{112}$$

where $[X]_t^c$ is the continuous part of the quadratic variation process of X, namely the non-decreasing process defined as

$$[X]_t^c = [X]_t - \sum_{0<s\leq t} (\Delta X_s)^2 - X_0^2 \tag{113}$$

$$= [X]_t - \sum_{0\leq s\leq t} (\Delta X_s)^2 \tag{114}$$

with

$$[X]_0^c = 0. \tag{115}$$

The interpretation is standard. Formula (112) represents the amount of time, in information units, that the càdlàg semimartingale spends in an arbitrarily small right neighborohood of a between time 0 and time T. Differently put, local time is an occupation density relative to the random clock $d[X]_s^c$.

A corresponding notion in calendar units can be easily obtained after noticing that $d[X]_t^c = \sigma^2(X_{t-})dt$ in the presence of a process whose dynamics are driven by (93) above. In fact,

$$\overline{L}_X(T,a) = \frac{1}{\sigma^2(a)} L_X(T,a) \tag{116}$$

is the chronological counterpart of (112) for SJDP's of the type analyzed here. As in the case of standard SDP's, the chronological sojourn time (116) can be intepreted as a version of the Radon-Nikodym derivative of the occupation measure with respect to the Lebesgue measure, i.e.

$$\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds = \int_A \overline{L}_X(T,a)da, \qquad \forall A \in \mathfrak{B}(\mathfrak{D}), \tag{117}$$

and, as pointed out above, is an occupation density. Additionally, formula (117) readily leads to

$$[X]_t^c = \int_{-\infty}^{\infty} L_X(t,a)da \tag{118}$$

which can be interpreted as a decomposition of variance (c.f. (29)), coherently with our remarks in Section 3.

We now turn to estimation. As earlier when dealing with SDP's, there is a natural way to identify (116) using a sample of observations generated from (93), namely we can perform density-like kernel estimation as in (35). Assume the same sampling scheme as in Section 2. The following result is proved in BN (2000).

Theorem 4.1 *Assume X_t is the solution to (93). If $h_{n,\overline{T}} \to 0$ as $n \to \infty$ with $T = \overline{T}$ in such a way that $\frac{1}{h_{n,\overline{T}}}(\Delta_{n,\overline{T}} \log(1/\Delta_{n,\overline{T}})^{1/2} = o(1)$, then*

$$\widehat{\overline{L}}_X(\overline{T},a) = \frac{\Delta_{n,\overline{T}}}{h_{n,\overline{T}}} \sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,\overline{T}}} - a}{h_{n,\overline{T}}}\right) \overset{a.s.}{\to} \overline{L}_X(\overline{T},a). \tag{119}$$

Proof *See BN (2000).*

Coherently with Thereom 3.1 in Section 3, Theorem 4.1 justifies using density-like kernel estimates as descriptive tools for SJDP's even in the presence of processes that might not possess a time-invariant stationary distribution. All we need to do is modify their intepretation since, in general, they cannot be regarded as estimates of a stationary density and recognize instead the role played by local time in characterizing the locational features of the process.

We conclude this subsection with two observations. First, recurrence implies divergence of the local time process as $T \to \infty$, just as when dealing with standard SDP's. In general, the rate of divergence cannot be quantified, although we expect positive recurrent and stationary processes to have local times that diverge at the fastest rate T. As in the case of the estimation of the infinitesimal moments of an SDP, the divergence properties of the local time factor affect the convergence properties of the infinitesimal moment estimators of (93) above, as shown in the next subsection. Second, functions of spatial densities, such as spatial hazard rates, can be readily defined as in Section 3. The intuition is immediate and follows our discussion in the previous section. We do not dwell on it here.

We now turn to the estimation of the infinitesimal moments and their potential role in the identification of general SJDP's.

4.2 Kernel Estimation of the Infinitesimal Moments of an SJDP

As pointed out earlier, identification of a recurrent solution to (93) above entails estimation of four quantities: the drift $\mu(.)$, the diffusive volatility $\sigma^2(.)$, the intensity of a jump $\lambda(.)$ and the distribution of the jump component. Such quantities can be backed out from the estimated infinitesimal moments as shown by the following example which is due to Johannes (2000).

Assume $g(x, y) = y$ where y is normally distributed with mean 0 and variance σ^2. Then, we can rewrite (109), (110) and (111) with $k = 4$ and $k = 6$ as

$$\begin{aligned}
\mathbf{M}^1(a) &= \mu(a), & (120)\\
\mathbf{M}^2(a) &= \sigma^2(a) + \lambda(a)\mathbf{E}_Y[y^2] = \sigma^2(a) + \lambda(a)\sigma_y^2, & (121)\\
\mathbf{M}^4(a) &= 3\left(\sigma_y^2\right)^2 \lambda(a), & (122)\\
\mathbf{M}^6(a) &= 15\left(\sigma_y^2\right)^3 \lambda(a), & (123)
\end{aligned}$$

$\forall a \in \mathfrak{D}$. Formula (120) through (123) suggest that the objects of interest can be derived as the solution to a system of four equations in four unknowns provided the moments are estimable. Following BP (1998), Johannes (2000) recommends using kernel counterparts to the infinitesimal moments as in (48) and (49) above to identify (120), (121), (122) and (123), namely

$$\widehat{\mathbf{M}}^k_{(n,T)}(a) = \frac{1}{\Delta_{n,T}} \frac{\sum_{i=1}^{n-1} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)\left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}\right)^k}{\sum_{i=1}^{n} \mathbf{K}\left(\frac{X_{i\Delta_{n,T}} - a}{h_{n,T}}\right)}, \tag{124}$$

$\forall k \geq 1$, where $\mathbf{K}$ is a standard kernel function (c.f. the assumptions in Section 2). Apparently, if $\widehat{\mathbf{M}}^k(a)$ is consistent for the theoretical k'th moment, then kernel estimation of the quantities of interest can be performed based on the following scheme (c.f. Johannes (2000)). First, obtain an estimate of σ_y^2 using

$$\left(\widehat{\sigma}_y^2\right)_{(n,T)} = \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{\mathbf{M}}^6_{(n,T)}(X_{i\Delta_{n,T}})}{5\widehat{\mathbf{M}}^4_{(n,T)}(X_{i\Delta_{n,T}})}. \tag{125}$$

Second, estimate $\lambda(a)$ via

$$\widehat{\lambda}_{(n,T)}(a) = \frac{\widehat{\mathbf{M}}^4_{(n,T)}(a)}{3\left(\widehat{\sigma}_y^4\right)_{(n,T)}}. \tag{126}$$

Third, estimate $\sigma^2(a)$ using

$$\widehat{\sigma}^2_{(n,T)}(a) = \widehat{\mathbf{M}}^2_{(n,T)}(a) - \widehat{\lambda}_{(n,T)}(a) \left(\widehat{\sigma}^2_y\right)_{(n,T)}. \tag{127}$$

Fourth, estimate $\mu(a)$ via

$$\widehat{\mathbf{M}}^1_{(n,T)}(a). \tag{128}$$

We assume the same sampling mechanism as in Section 2. The following theorem shows almost sure consistency of $\widehat{\mathbf{M}}^k_{(n,T)}(a)$ for $\mathbf{M}^k(a)$ $\forall k \geq 1$, thereby justifying the scheme that we laid out previously, and it provides an asymptotic distribution.

Theorem 4.3 *Assume X is the solution to (93). If $n \to \infty$, $T \to \infty$, $\frac{T}{n} \to 0$ and $h_{n,T} \to 0$ so that $\frac{\overline{L}_X(T,a)}{h_{n,T}}(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2} = o_{a.s.}(1)$ and $h_{n,T}\overline{L}_X(T,a) \stackrel{a.s.}{\to} \infty$ $\forall a \in \mathfrak{D}$, then*

$$\widehat{\mathbf{M}}^k_{(n,T)}(a) \stackrel{a.s.}{\to} \mathbf{M}^k(a) \qquad \forall k \geq 1. \tag{129}$$

Furthermore, if $h^5_{n,T}\overline{L}_X(T,a) = O_{a.s.}(1)$ $\forall a \in \mathfrak{D}$, then

$$\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}\left(\widehat{\mathbf{M}}^k_{(n,T)}(a) - \mathbf{M}^k(a) - \Gamma_{\mathbf{M}^k}(a)\right) \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2\left(\mathbf{M}^{2k}(a)\right)\right), \tag{130}$$

$\forall k \geq 1$, where

$$\Gamma_{\mathbf{M}^k}(a) = h^2_{n,T}\mathbf{K}_1\left(\frac{1}{2}\left(\mathbf{M}^k(a)\right)^{''} + \left(\mathbf{M}^k(a)\right)^{'}\frac{m^{'}(a)}{m(a)}\right) \qquad \forall k \geq 1, \tag{131}$$

$m(dx)$ is the invariant measure of the process, $\mathbf{K}_1 = \int_{\mathfrak{R}} s^2\mathbf{K}(s)ds$ and $\mathbf{K}_2 = \int_{\mathfrak{R}} \mathbf{K}^2(s)ds$.

Proof *See BN (2000).*

Some observations are in order. First, simple parametric assumptions appear to be necessary for identification. Previously, we assumed a stylized structure for the jump impact $g(.,y)$, namely we imposed $g(.,y) = y$ with y normally distributed. Furthermore, a straighforward mechanism was put forward to estimate the functions and parameters of interest. A more complicated methodology that uses moments of order higher than six (and, consequently, takes into account their information content) could have been suggested instead. Clearly, any identification scheme entails averages of nonparametric estimates for the purpose of the estimation of the parameters of the model (c.f. (125) above). Hence, while the estimates of the functions of interest (viz. $\mu(.)$, $\sigma^2(.)$ and $\lambda(.)$) converge at the nonparametric rate $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$ (by a simple application of the delta method), the parameter estimate $\widehat{\sigma}^2_y$ is expected to converge at a faster rate. In other words, while causing a loss in terms of generality of the model, the necessary (for identification) imposition of parametric assumptions is, not surprisingly, beneficial for estimation.

Second, the methodology is rather flexible. Different parametric structures and identification schemes could have been described. In addition, the functions $\mu(.)$, $\sigma^2(.)$ and $\lambda(.)$ are generally unrestricted, thereby allowing for non-linearities (i.e. non-affine structures) that might be particularly appealing in some continuous-time finance models. We refer the interested reader to Johannes (2000), where (93) and the above identification scheme are applied to the study of the short-term interest rate process.

In virtue of our discussion in the pure diffusion case, the implications of Theorem 4.3 should be clear. Here we focus on the main differences between the case with discontinuities and the case without discontinuities examined in the previous section. Contrary to diffusion estimation (c.f. Section 3), all the infinitesimal moment estimators converge at the same rate, namely $\sqrt{h_{n,T}\widehat{\overline{L}}_X(T,a)}$. The intuition is as follows. The estimator (124) above hinges on averages of terms of order $\sqrt{dt}$ for every $k \geq 1$. In the case of standard SDP's, the drift estimator (48) is an average of terms of order $\sqrt{dt}$, whereas the diffusion estimator (49) averages terms of order dt, leading to a faster rate of convergence for the latter. Apparently, no infinitesimal moments can be identified over a fixed span of data in the presence of an underlying SJDP. In fact, if T were fixed, then $\widehat{\overline{L}}_X(T,a)$ would be bounded in probability and $\sqrt{h_{n,T}}$ would not be a proper rate of convergence. Again, the intuition is simple. In jump-diffusion models like (93) above, the function $\mu(.)$ has the same interpretation as in the standard set-up without jumps. As a consequence, an enlarging span of data is expected to be necessary for the consistency of $\widehat{\mathbf{M}}^1_{(n,T)}(.)$. As for the higher moments, formulae (109), (110) and (111) illustrate their dependence on the characteristics of the jump component. It is known that a fixed span of data does not contain sufficient information for the identification of the features of the jump part.

5 Multivariate Diffusion Processes (MDP's)[8]

In this section we model a time-series as the d-dimensional solution X_t of the multivariate stochastic differential equation

$$dX_t = \boldsymbol{\mu}\left(X_t\right) dt + \boldsymbol{\sigma}\left(X_t\right) d\mathbf{B}_s \tag{132}$$

where $\mathbf{B} = \left\{\mathbf{B}_t, \Im_t^B; 0 \leq t < \infty\right\}$ is an m-dimensional standard Brownian motion, $\boldsymbol{\mu}(.) = \{\mu_i(.)\}_{1\leq i\leq d}$ is a $d \times 1$ Borel measurable vector, $\boldsymbol{\sigma}(.) = \{\sigma_{ij}(.)\}_{\substack{1\leq i\leq d \\ 1\leq j\leq m}}$ is a $d \times m$ Borel measurable matrix and $X_0 = \overline{X} \in \mathfrak{D} \subseteq \Re^d$ is a given initial condition that is taken to be independent of $\Im_\infty^B$ and with finite second moment, i.e. $\mathbf{E}||\overline{X}|| < \infty$. We define the left-continuous filtration

[8] This section is a largely abridged and simplified version of BM (2001). We refer the interested reader to that paper for additional discussions, technical details and proofs.

$$\overline{\Im}_t := \sigma(\overline{X}) \vee \Im_t^B = \sigma(\overline{X}, \mathbf{B}_s; 0 \leq s \leq t) \qquad 0 \leq t < \infty \tag{133}$$

and the collection of null sets

$$\aleph := \{N \subseteq \Omega; \exists G \in \overline{\Im}_\infty \textit{ with } N \subseteq G \textit{ and } P(G) = 0\}. \tag{134}$$

We then create the *augmented* filtration

$$\widetilde{\Im}_t^X := \sigma(\overline{\Im}_t \cup \aleph) \qquad 0 \leq t < \infty. \tag{135}$$

Assumption 1 and 2 (1 and 2bis) below are imposed on (132) to guarantee the existence of a unique and null-recurrent (positive recurrent) solution to (132).

(1) *$\boldsymbol{\mu}(.)$ and $\boldsymbol{\sigma}(.)$ are time-homogeneous, $\mathfrak{B}$-measurable functions on $\mathfrak{D} \subseteq \Re^d$ where $\mathfrak{B}$ is the σ-field generated by Borel sets on $\mathfrak{D}$. Both functions satisfy local Lipschitz and linear growth conditions. Thus, for every compact subset J of the range of the process, there exist constants C_1 and C_2 such that, for all x and y in J,*

$$||\boldsymbol{\mu}(x) - \boldsymbol{\mu}(y)|| + ||\boldsymbol{\sigma}(x) - \boldsymbol{\sigma}(y)|| \leq C_1||x - y||, \tag{136}$$

and

$$||\boldsymbol{\mu}(x)|| + ||\boldsymbol{\sigma}(x)|| \leq C_2\{1 + ||x||\}, \tag{137}$$

where $||\boldsymbol{\sigma}|| = \sum_{i=1}^d \sum_{i=1}^m \sigma_{ij}^2$ and $||\boldsymbol{\mu}|| = \sum_{i=1}^d \mu_i^2$.

(2) *(Null recurrence) Define the positive definite matrix $\mathbf{s}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)'$ such that $s_{ik}(x) = \sum_{g=1}^m \sigma_{ig}(x)\sigma_{gk}(x)$ $\forall x \in \mathfrak{D} \subset \Re^d$ and assume that every open and bounded set $A \in \mathfrak{D}$ satisfies*

$$\min_{x \in \overline{A}} s_{ii}(x) > 0, \tag{138}$$

for some $1 \leq i \leq d$. Now define the second-order elliptic operator

$$\mathfrak{L}\varphi(.) = \sum_{i=1}^d \mu_i(.)\frac{\partial\varphi(.)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^d\sum_{k=1}^d s_{ik}(x)\frac{\partial\varphi(.)}{\partial x_i \partial x_k}. \tag{139}$$

There is a function $\varphi(.) : \Re^d \backslash \{0\} \to \Re$ of class C^2 in the domain of the operator that satisfies

$$\mathfrak{L}\varphi(.) \leq 0 \qquad \textit{on } \Re^d \backslash \{0\}, \tag{140}$$

and is such that $\Psi(r) := \min_{||x||=r} \varphi(.)$ is strictly increasing with $\lim_{r\to\infty} \Psi(r) = \infty$ (c.f. Karatzas and Shreve, 1991, Exercise 7.13, part (i), page 370).

(2bis) *(Positive recurrence) There is a function* $\varphi(.) : \Re^d \backslash \{0\} \rightarrow \Re$ *of class* C^2 *in the domain of the operator that satisfies*

$$\mathfrak{L}\varphi(.) \leq -1 \qquad on\ \Re^d \backslash \{0\}, \tag{141}$$

and is such that $\Psi(r) := \min_{||x||=r} \varphi(.)$ *is strictly increasing with* $\lim_{r\to\infty} \Psi(r) = \infty$ *(c.f. Karatzas and Shreve, 1991, Exercise 7.13, pat (iii), page 371).*

Under Assumptions 1 and 2 (2 bis), the stochastic differential equation (132) displays a strong solution X_t that is unique and null recurrent (positive recurrent). Specifically, the process X_t satisfies

$$X_t = \overline{X} + \int_0^t \boldsymbol{\mu}(X_s)\, ds + \int_0^t \boldsymbol{\sigma}(X_s)\, d\mathbf{B}_s, \tag{142}$$

and is square integrable, i.e. $\mathbf{E}||X_t||^2 < \infty \quad \forall t$. Equivalently, we can write each coordinate X_t^j as

$$X_t^j = \overline{X}^j + \int_0^t \mu_j(X_s)\, ds + \sum_{g=1}^m \int_0^t \sigma_{jg}(X_s)\, dB_s^g, \qquad 0 \leq t < \infty, 1 \leq j \leq d. \tag{143}$$

Coherently with the scalar model in Section 3, the dynamics of X_t are determined by Brownian shocks and by the functional forms of the matrices $\boldsymbol{\mu}(.)$ and $\mathbf{s}(.)$ (recall from Assumption (2) above that $\mathbf{s}(x) = \boldsymbol{\sigma}(x)\boldsymbol{\sigma}(x)'$). Such matrices will be the object of econometric interest in the present section. As in the scalar case, they both have straightforward representations in terms of infinitesimal conditional moments. In particular,

$$\mathbf{E}^a\left[X_s^i - a_i\right] = t\mu_i(a) + o(t) \tag{144}$$

$$\mathbf{E}^a\left[\left(X_s^i - a_i\right)\left(X_s^j - a_j\right)\right] = ts_{ij}(a) + o(t) \tag{145}$$

as $t \downarrow 0$ (c.f. Karatzas and Shreve (1991)).

The notions of recurrence that Assumption 2 implies are standard (c.f. Section 2). Namely, the multidimensional process X_t is Harris recurrent if there is a $\sigma-$finite measure $m(dx)$ such that $m(A) > 0$ implies $\lim_{T\to\infty} \eta_A^T = \infty$ with probability one $\forall A \in \mathfrak{B}(\mathfrak{D})$ where $\eta_A^T = \int_0^T \mathbf{1}_{\{X_s \in A\}} ds$ is, as earlier, the occupation time measure of A. The following result gives us the rate at which η_A^T diverges to infinity and a weak convergence result for η_A^T, inter alia.

Theorem 5.1 *Assume* X_t *is the solution to (132). Consider the non-negative function* $\delta(.)$. *If there exists a constant* $\alpha \in [0, 1]$ *and a slowly varying function at infinity* $L(T)$ *such that*

$$\lim_{T\to\infty} \mathbf{E}_x \left[\int_0^T \delta(X_s)ds\right] / T^\alpha L(T) = C_X > 0 \qquad \forall x \in \mathfrak{D}, \tag{146}$$

then

$$\lim_{T\to\infty} \Pr\left\{\frac{1}{C_X u(T)} \int_0^T \delta(X_s)ds < x\right\} = G_\alpha(x), \tag{147}$$

where

$$G_\alpha(x) = \frac{1}{\pi\alpha} \int_0^x \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \sin \pi\alpha j \Gamma(\alpha j + 1) y^{j-1} dy, \tag{148}$$

$u(T) = T^\alpha L(T)$, $\Gamma(.)$ *is the Gamma function,* $C_X = C_X^* \int_{-\infty}^{\infty} \delta(x)m(dx)$, $m(dx)$ *is the invariant measure and* C_X^* *is a process-specific constant.*

Proof *See Darling and Kac (1956) for the original statement of the theorem. Bingham (1971) contains a functional version of the same finding. Several papers discuss limit results for slowly increasing occupation times associated with null-recurrent Markov processes, see Höpfner and Löcherbach (2001) for a complete recent survey of the literature on the subject. The interested reader is referred to Khasminskii (2001) and Khasminskii and Yin (2000) for a detailed treatment of the one-dimensional null-recurrent diffusion case.*

We can rewrite (147) as follows:

$$\frac{\int_0^T \delta(X_s)ds}{u(T)} \Rightarrow C_X g_\alpha, \tag{149}$$

where g_α is the Mittag-Leffler distribution with parameter α, i.e.

$$g_\alpha(x) = \frac{1}{\pi\alpha} \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \sin \pi\alpha j \Gamma(\alpha j + 1) x^{j-1}. \tag{150}$$

Theorem 5.1 shows that additive functionals of the process (c.f., $\int_0^T \delta(X_s)ds$) converge weakly (in distribution) to a random variable endowed with the Mittag-Leffler density g_α when standardized appropriately (by $u(T)$). The rate of divergence to infinity of the standarding factor (and, as a consequence, the rate of divergence to infinity of the continuous averages) depends on the statistical features of the process through the constant α (c.f. $u(T) = L(T)T^\alpha$). Naturally, α affects the features of the Mittag-Leffler distribution as well.

Some observations are in order. First, the theorem readily applies to the occupation measures in the definition of recurrence since we can take $\delta(.) = \mathbf{1}(.)$ giving $\int_0^T \delta(X_s)ds = \eta_A^T$ $\forall A \in \mathfrak{B}(\mathfrak{D})$. Second, it extends to most continuous-time Harris recurrent Markov processes, provided assumption (146) is satisfied (c.f. Darling and Kac (1956)). In fact, in general, we could apply it to SDP's and SDJP's of the type analyzed in the present review. Nonetheless, its implications appear to be particularly interesting when analyzing processes for which a standard notion of local time cannot be defined, as is the case with multivariate diffusions. We do not dwell on this idea here (and refer the reader to BM (2001)) but the meaning of the statement will become clear in the next subsection.

We now briefly consider some interesting special cases. We noticed earlier that the characteristics of the underlying recurrent process affect the weak convergence result through the constant α which modifies both the rate of divergence of the occupation measure and the limiting distribution. The constant is known for few processes. Specifically, if X_t is Brownian motion on the plane, then $\alpha = 0$ and the Mittag-Leffler distribution coincides with the exponential distribution, i.e.

$$\frac{\int_0^T \delta(X_s)ds}{\log T} \Rightarrow C_X e^{-x} \qquad with \quad x \geq 0. \tag{151}$$

If X_t is a scalar Brownian motion, then $\alpha = \frac{1}{2}$ and the Mittag-Leffler density is equal to the truncated normal, i.e.

$$\frac{\int_0^T \delta(X_s)ds}{\sqrt{T}} \Rightarrow C_X \frac{1}{\sqrt{\pi}} e^{-x^2/4} \qquad with \quad x \geq 0, \tag{152}$$

implying that the dimensionality of the system has, in general, an impact on the rate of divergence of the continuous averages. In effect, we go from a $\sqrt{T}$-rate to a $\log T$-rate when moving from the scalar Brownian motion case to its bivariate counterpart. Interestingly, the dimensionality of the process does not influence the rate of divergence of the continuous averages (or the rate of divergence of the occupation measures) if stationarity is satisfied. Under stationarity $\alpha = 1$ and

$$\frac{\int_0^T \delta(X_s)ds}{T} \xrightarrow{p} \int_{-\infty}^{\infty} \delta(x) f(dx), \tag{153}$$

where $f(dx)$ is the time-invariant stationary probability measure of X_t, which is a form of the classical ergodic theorem. We will return to Theorem 5.1 in the sequel. We now consider generalized density estimation for multivariate solutions to (132).

5.1 Generalized Density Estimation for MDP's

Just as it seems natural to estimate multivariate density functions using multidimensional extensions of kernel estimates for scalar densities (c.f. Pagan and Ullah (1999)), it might appear natural to estimate the local time of a vector process using a multivariate counterpart of the standard estimator from Section 3, i.e.

$$\widehat{\widehat{\mathbb{L}}}_{(n,T)}(\overline{T}, a) = \frac{\Delta_{n,\overline{T}}}{\mathbf{h}_{n,\overline{T}}} \sum_{i=1}^{n} \left(\prod_{j=1}^{d} \mathbf{K} \left(\frac{X^j_{i\Delta_{n,\overline{T}}} - a_j}{h_{n,\overline{T}}} \right) \right), \tag{154}$$

where $\mathbf{h}_{n,\overline{T}} = h^d_{n,\overline{T}}$ and $a = (a_1, a_2, .., a_d) \in \Re^d$. As it happens, local time is not generally defined for multidimensional semimartingales (c.f. Brugière (1991), *inter alia*). In consequence, we cannot build a notion of (spatial) density for multivariate, potentially nonstationary, continuous-time processes based on local time as suggested in Section 3 for SDP's and in Section 4 for SJDP's. Consistently, over a fixed span of data $\overline{T}$, the quantity $\widehat{\overline{\mathbb{L}}}_{(n,\overline{T})}(\overline{T}, a)$ cannot be interpreted as a multivariate sojourn time estimator despite being a local time estimator for $d = 1$. Nonetheless, its asymptotic features as $n \to \infty$ for a fixed $T = \overline{T}$ can still be characterized. Using a multivariate indicator kernel (but we expect the results not to change in the presence of a continuous kernel function), Brugière (1993) shows that

$$\frac{1}{\mathbf{h}_{n,\overline{T}} \log\left(1/\mathbf{h}^2_{n,\overline{T}}\right)} \widehat{\overline{\mathbb{L}}}_{(n,\overline{T})}(\overline{T}, a) \tag{155}$$

converges weakly (as $n \to \infty$) to an exponentially distributed random variable when the dimension of the system d is equal to two. Furthermore, the quantity

$$\frac{1}{\mathbf{h}_{n,\overline{T}}} \widehat{\overline{\mathbb{L}}}_{(n,\overline{T})}(\overline{T}, a) \tag{156}$$

converges weakly (in distribution) to

$$\int_0^\infty \mathbf{1}_{\{\sigma(0)\mathbf{B}_s < 1\}} ds \tag{157}$$

when $d \geq 3$.

Interestingly, while preventing us from constructing appealing descriptive statistics for multidimensional (potentially nonstationary) semimartingales based on (154), the nonexistence of a notion of local time is not prohibitive when it comes to dealing with the estimation of the infinitesimal moments of (132). This result might at first appear surprising since the local time estimates are known to play a fundamental role in the scalar limit theory for recurrent SDP's and SJDP's. On the other hand, it is well know that simple matrix functionals of multivariate processes, such as the functional $\int_0^t \mathbf{B}(s) \mathbf{B}(s)' ds$ of the vector Brownian motion B, are well defined and sample functions converge weakly to them, even though these functionals may not have a representation in terms of local time as they do from the occupation formula in the scalar case.

Coherently, we now discuss a finding that is crucial in building an estimation theory for recurrent, multivariate, diffusions without resorting to a notion of local time. The following result, which heavily hinges on Thereom 5.1. above, characterizes the behavior of $\widehat{\overline{\mathbb{L}}}_{(n,T)}(T, a)$ over an enlarging span of observations, that is as $T \to \infty$ (with $n \to \infty$). In the next subsection we will show that the asymptotic behavior of $\widehat{\overline{\mathbb{L}}}_{(n,T)}(T, a)$ as $T, n \to \infty$ is crucial in interpreting the limit theory of the kernel estimates of the infinitesimal moments of the solution to (132). We use the sampling scheme that was laid out in Section 2.

Theorem 5.2 *Assume X_t is the solution to (132) and (146) is satisfied. If the bandwidth $h_{n,T}$ is such that*

$$\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a) \stackrel{a.s.}{\to} \infty \tag{158}$$

and

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}}{\mathbf{h}_{n,T}}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a) = o_{a.s.}(1) \tag{159}$$

as $h_{n,T} \to 0$ with $n, T \to \infty$ so that $\Delta_{n,T} \to 0$ $\forall a \in \mathfrak{D} \subseteq \mathfrak{R}^d$, then

$$\frac{\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{u(T)} \Rightarrow C_X m\,(a)\,g_\alpha \quad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^d \tag{160}$$

for some function $u(T)$ which is regularly varying at infinity with parameter α satisfying $0 \leq \alpha \leq 1$, i.e.

$$u(T) = L(T)T^\alpha, \tag{161}$$

with $L(T)$ slowly varying, where g_α is the Mittag-Leffler distribution with the same parameter α and $m(dx)$ is the invariant measure of the process. C_X is a process specific constant.

Proof *See BM (2001).*

Theorem 5.2 links the divergence properties of $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ to those of the occupation time measure η_A^T. This result is hardly surprising, being that $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ is an estimate of the time spent by the process in the vicinity of the spatial point a even though the dimensionality of the system prevents us from intepreting it as a consistent estimate of the local time of the process at a.

Two observations are in order. First, Theorem 5.2 applies to SDP's of the type analyzed in Section 3. Previously, we pointed out that the local time estimates of stationary processes and standard scalar Brownian motion diverge at rate T and $\sqrt{T}$, respectively. The same result can be deduced from Theorem 5.2 as a subcase of the more general theory laid out in this section. Second, in the presence of stationary processes of any dimension $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ represents a well-defined density estimator. In fact, if $\alpha = 1$, then $g_\alpha = 1$ and $C_X = \frac{1}{m(\mathfrak{D})}$. Hence,

$$\frac{\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{T} = \frac{1}{n\mathbf{h}_{n,T}}\sum_{i=1}^{n}\left(\prod_{j=1}^{d}\mathbf{K}\left(\frac{X^j_{i\Delta_{n,T}} - a_j}{h_{n,T}}\right)\right) \tag{162}$$

converges to $\frac{m(a)}{m(\mathfrak{D})} = f(a)$, which is a standard finding in multivariate density estimation for both discrete and continuous-time stationary processes (c.f. Prakasa-Rao (1983) and Silverman (1986)). We now turn to the estimation of the infinitesimal moments.

5.2 Kernel Estimation of the Infinitesimal Moments of an MDP

Following our discussion in the previous sections, it is natural to estimate the matrices $\boldsymbol{\mu}(.)$ and $\mathbf{s}(.) = \boldsymbol{\sigma}(.)\boldsymbol{\sigma}(.)'$ using nonparametric kernel estimates of the Nadaraya-Watson type, i.e.

$$\widehat{\boldsymbol{\mu}}_{(n,T)}(a) = \frac{\frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}_i^a \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right)}{\sum_{i=1}^{n} \mathbf{K}_i^a}, \tag{163}$$

and

$$\widehat{\mathbf{s}}_{(n,T)}(a) = \frac{\frac{1}{\Delta_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}_i^a \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right) \left(X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}} \right)'}{\sum_{i=1}^{n} \mathbf{K}_i^a} \tag{164}$$

where

$$\mathbf{K}_i^a = \prod_{j=1}^{d} \mathbf{K} \left(\frac{X_{i\Delta_{n,T}}^j - a_j}{h_{n,T}} \right). \tag{165}$$

As pointed out earlier, the nonexistence of local time for multivariate solutions to (132) above does not represent an impossible obstacle when deriving an estimation theory based on (163) and (164). The intuition relies on the following observations: any limit results for (163) and (164) in the $d \geq 1$ case should collapse in the findings that we illustrated in Section 3, i.e. (51) and (54), when reducing the dimensionality of the system to $d = 1$. Let us focus on the drift for illustration purposes. Based on (51), our best guess of a weak convergence result for (163) is:

$$\sqrt{\mathbf{h}_{n,T} \widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)} \left\{ \widehat{\boldsymbol{\mu}}_{(n,T)}(a) - \boldsymbol{\mu}(a) \right\} \Rightarrow \mathbf{N}\left(0, \mathbf{K}_2 \mathbf{s}(a)\right) \tag{166}$$

where $\mathbf{h}_{n,T} = h_{n,T}^d$. Two observations are in order. First, (166) reduces to (51) when $d = 1$, thereby satisfying our requirement. Second, the impossibility of interpreting $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ as a local time estimator for $d > 1$ does not have an impact on the credibility of the intuition leading to (166). In fact, as shown earlier, $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ converges (as $n, T \to \infty$ and if standardized appropriately) to a well-defined random variable for dimensions higher than one while also being a local time estimator when $d = 1$. The following theorem confirms this intuition. As usual, we adopt the same sampling scheme as in Section 2.

Theorem 5.3 *Assume X_t is the solution to (132) and (146) is satisfied. If*

$$\frac{(\Delta_{n,T} \log(1/\Delta_{n,T}))^{1/2} \widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{\mathbf{h}_{n,T}} = o_{a.s.}(1) \tag{167}$$

and

$$\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)h_{n,T}^{d} \overset{a.s.}{\to} \infty \tag{168}$$

as $h_{n,T} \to 0$ *with* $n, T \to \infty$ *and* $\Delta_{n,T} \to 0$, *then*

$$\widehat{\boldsymbol{\mu}}_{(n,T)}(a) \overset{a.s.}{\to} \boldsymbol{\mu}(a) \qquad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^{d}. \tag{169}$$

Furthermore, if $h_{n,T} = O_{a.s.}\left(\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)^{-\frac{1}{d+4}}\right)$, *then*

$$\begin{aligned} &\sqrt{\mathbf{h}_{n,T}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}\left(\widehat{\boldsymbol{\mu}}_{(n,T)}(a) - \boldsymbol{\mu}(a) - \boldsymbol{\Gamma}_{\boldsymbol{\mu}}(a)\right) \\ \Rightarrow \quad &(\mathbf{s}(a))^{1/2}\,\mathbf{N}\left(0, \mathbf{K}_2\mathbf{I}\right) \qquad\qquad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^{d} \end{aligned} \tag{170}$$

where

$$\boldsymbol{\Gamma}_{\boldsymbol{\mu}}(a) = (bias_1, bias_2, ..., bias_d)(a), \tag{171}$$

$$bias_i(a) \quad = \quad h_{n,T}^{2}\mathbf{K}_1\left(\sum_{k=1}^{d}\frac{\partial\mu_i\left(a\right)}{\partial a_k}\frac{\frac{\partial m(a)}{\partial a_k}}{m(a)} + \frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2\mu_i\left(a\right)}{\partial a_k\partial a_k}\right) \qquad \forall i = 1, ..., d, \tag{172}$$

$m\left(dx\right)$ *is the invariant measure of the process,* $\mathbf{K}_1 = \int s^2\mathbf{K}(s)ds$ *and* $\mathbf{K}_2 = \left(\int \mathbf{K}^2(s)ds\right)^d$.

Proof *See BM (2001).*

All our comments in the scalar case apply to the multivariate set-up examined here up to some minor modifications. We will therefore not be as detailed as in Section 3. Nonetheless, it should be noted that the asymptotic bias is $O(h_{n,T}^2)$, as in the scalar case, whereas the asymptotic variance is of order $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)^{-1}h_{n,T}^{-d}$ rather than $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)^{-1}h_{n,T}$. In the standard estimation of conditional first moments in the discrete-time, stationary, framework, the limiting bias is $O(h_{n,T}^2)$ while the limiting variance is $n^{-1}h_{n,T}^{-d}$, rather than $(nh_{n,T})^{-1}$. In other words, the variance increases with the dimensionality of the system. This effect is commonly known as *the curse of dimensionality* (c.f. Silverman (1986)). Here we have a curse of dimensionality that mirrors the classical result in conventional nonparametric estimation of conditional moments in discrete time and manifests itself through the factor h^{-d} (i.e. the bandwidth sequence that inversely affect the order of the variance term is raised to a power that increases with the dimensionality of the system), as well as an additional curse of dimensionality that operates via the quantity $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$. The latter effect is a genuine by-product of the generality of this theory and, in particular, is due to the robustness to deviations from stationarity. In fact, should the system be stationary (or positive recurrent), then $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ would diverge at speed T (c.f. the previous subsection) regardless of the number of equations and the order of the variance term would simply be

$Th^d_{n,T}$. Hence, we would be in the presence of the classical dimensionality problem since only the power d would be affected by the number of equations in the system. Now, consider the null recurrent situation. We pointed out earlier that scalar Brownian motion and Brownian motion on the plane imply divergence rates for $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ that are equal to $\sqrt{T}$ and $\log T$, respectively (see the previous subsection). This result has broader implications. We expect the dimensionality of the system to have a negative influence on the rate of divergence of the factor $\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)$ for null recurrent processes that are more general than Brownian motion, thereby reinforcing the conventional effect that comes into play through the term $h^d_{n,T}$ and leading to a slower rate of convergence of the nonparametric estimates to the theoretical vector $\boldsymbol{\mu}(.)$. The optimal bandwidth sequence, i.e.

$$h_{n,T} \propto \widehat{\widehat{\mathbb{L}}}^{-\frac{1}{d+4}}_{(n,T)}(T,a)$$

accounts for both effects (i.e. the *two curses of dimensionality* in the terminology of BM (2001)).

We now turn to diffusion estimation. The symbol $\otimes$ in the statement of Theorem 5.4 stands for the standard Kronecker product.

Theorem 5.4 *Assume X_t is the solution to (132) and (146) is satisfied. If*

$$\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a) \overset{a.s.}{\to} \infty \tag{173}$$

and

$$\frac{(\Delta_{n,T}\log(1/\Delta_{n,T}))^{1/2}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{\mathbf{h}_{n,T}} = o_{a.s.}(1) \tag{174}$$

as $h_{n,T} \to 0$ with $n,T \to \infty$ and $\Delta_{n,T} \to 0$, then

$$\widehat{\mathbf{s}}_{(n,T)}(a) \overset{a.s.}{\to} \mathbf{s}(a) \qquad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^d. \tag{175}$$

Furthermore, if $h^2_{n,T}\sqrt{\frac{\mathbf{h}_{n,T}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{\Delta_{n,T}}} = O_{a.s.}(1)$, *then*

$$\begin{aligned} &\sqrt{\frac{\mathbf{h}_{n,T}\widehat{\widehat{\mathbb{L}}}_{(n,T)}(T,a)}{\Delta_{n,T}}}\left(\text{vech}\widehat{\mathbf{s}}_{(n,T)}(a) - \text{vech}\mathbf{s}(a) - \mathbf{\Gamma}_{\boldsymbol{\sigma}^2}(a)\right) \\ \Rightarrow \quad & (\mathbf{\Xi}(a))^{1/2}\,\mathbf{N}\left(0,\mathbf{K}_2\mathbf{I}\right), \qquad\qquad \forall a \in \mathfrak{D} \subseteq \mathfrak{R}^d, \end{aligned} \tag{176}$$

where

$$\mathbf{\Gamma}_{\boldsymbol{\sigma}^2}(a) = (bias_{1,1}, bias_{2,1}, ..., bias_{d,d})(a), \tag{177}$$

$$bias_{i,j}(a) \;=\; h^2_{n,T}\mathbf{K}_1\left(\sum_{k=1}^{d}\frac{\partial\sigma_{i,j}(a)}{\partial a_k}\frac{\frac{\partial m(a)}{\partial a_k}}{m(a)} + \frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2\sigma_{i,j}(a)}{\partial a_k\partial a_k}\right) \qquad \forall i,j = 1,...,d, \tag{178}$$

$$\boldsymbol{\Xi}(a) = L_D \left(2\mathbf{s}(a) \otimes 2\mathbf{s}(a)\right) L_D', \tag{179}$$

$$L_D = (D^{'} D)^{-1} D^{'}, \tag{180}$$

D is the standard duplication matrix, i.e. the unique $d^2 \times (d(d+1))/2$ matrix such that L_D eliminates redundant elements, viz.,

$$\text{vechs}(a) = L_D \text{vecs}(a) = \begin{bmatrix} s_{1,1} \\ s_{2,1} \\ s_{2,2} \\ s_{3,1} \\ \cdots \\ s_{d,d} \end{bmatrix}, \tag{181}$$

$m(dx)$ is the invariant measure of the process, $\mathbf{K}_1 = \int s^2 \mathbf{K}(s) ds$ and $\mathbf{K}_2 = \left(\int \mathbf{K}^2(s) ds\right)^d$.

Proof *See BM (2001).*

Our comments in the scalar diffusion case (c.f. Section 3) and in the multivariate drift case should suffice to interpret the results in Theorem 5.4 above. Here we note that, as with the standard scalar diffusion case, the local properties of the process contain sufficient information to identify the diffusion matrix, i.e. $\mathbf{s}(a)$ can be estimated consistently over a fixed span of data $T = \overline{T}$. The interested reader is referred to the work of Brugiére (c.f. Brugière (1991, 1993)) for a thorough treatment in the fixed T case. In particular, Brugière (1991) discusses weak consistency of (164) for the matrix of interest, while Brugière (1993) proves the asymptotic normality of the diffusion matrix estimator. The kernel used in both papers is the discontinuous indicator kernel. Extension of the results in Brugière to the use of continuous kernels is immediate.

6 Application: the short-end of the term structure of interest rates

The recent empirical finance literature has devoted a substantial amount of attention to the short-term interest rate process. Research has mainly focused on the continuous-time specification of the spot rate dynamics for the purpose of fixed-income derivative pricing (c.f. Aït–Sahalia (1996), Bandi (1998), Conley *et al.* (1998) and the references therein, for example) and, following Fama and Schwert (1977), on the ability of the spot rate to predict stock returns at short and long horizons (c.f. Torous *et al.* (2001) and the references therein).

While standard preliminary nonstationarity tests support the view that the stationarity properties of commonly employed short-term interest rate series are unclear, the consequences of misattribution of stationarity for statistical inference in both discrete and continuous-time are likely to be non-negligible. Not surprisingly, some recent work on stock return predictability has resorted to near unit root specifications for the short rate when utilizing the short rate as a predictor for future stock returns (c.f. Torous *et al.* (2001)). In effect, while it is recognized that the conventional tests for (non)stationarity have low power against the close-to-unity alternative, the usual failure to reject the null hypothesis of nonstationarity would suggest explicit nonstationary behavior that some audience might find hard to fully accept on purely economic grounds.

The assumption of recurrence offers an alternative way to tackle this problem (c.f. Bandi (1998)). Short term rates might not possess a time-invariant stationary density as suggested but the unclear outcome of standard a-priori testing procedures. Nonetheless, it is natural to believe that their dynamic structures are such that they will return to values in their range with probability one as time elapses.

In this section we examine the stationarity properties of a selection of interest rate series by implementing a new testing procedure recently proposed by Moon and Perron (2001) and Phillips and Sul (2001). Unit root testing is conducted in an asymptotically large n and T panel endowed with an AR(1) specification, a common autoregressive parameter ρ and cross-sectional correlation induced by a factor structure. Conditional on the factors, the errors are assumed to be independent cross-sectionally (although they can be correlated in the time direction). Only a single factor is considered in Phillips and Sul (2001) work, whereas in Moon and Perron (2001) the number of factors is estimated using BIC and an alternative procedure recently discussed by Bai and Ng (2001) (called IC below). We follow the latter approach here and implement two classes of tests. The first is based on the raw data (level test), the second is based on the demeaned data (demeaned test). In both cases, the critical values for the t-statistics are based on an asymptotic standard normal distribution under the unit-root null hypothesis (i.e., $\rho = 1$). Rejection at the 5% level occurs in correspondence with t-statistics that are less than -1.64, as implied by the one-sided nature of the test. We refer the reader to Moon and Perron (2001) for additional details. Here it suffices to add that the test compares favorably to existing procedures in the presence of a large number of series.[9]

Initially, we employ a panel of 12 series that includes long term rates to obtain good power and size properties, under the (possibly unrealistic) maintained assumption of a common autoregressive parameter across rates with different maturities. Specifically, we use 1-month, 3-month and 6-month eurodollar rates, the federal fund rate, 3-month and 6 month t-bill rates, and constant-maturity bond rates between 1 and 10 years (namely, 1, 2,

[9]We thank Benoit Perron for help with the implementation of the test and extensive discussions.

3, 5, 7 and 10 years).[10] For each series we have a total of 6,418 observations from June 1, 1967 to March 15, 2002.

The results of the level test are in Table 1. Table 2 contains the outcome of the demeaned test. The former suggests rejection of the unit root null at the 5% level for any number of factors. The power is considerable given the high value of the estimated autoregressive parameter. The demeaned test provides conflicting results but appears to be slightly less reliable in virtue of its inferior performance in simulations (c.f. Moon and Perron (2001)).

One issue with tests based on panels is that it is not clear which series (if any) drive the results in practise. In our case we expect the exclusion of the longer rates to have the potential for reducing the extent of the rejection in the case of the level test. In light of this observation, we now simply focus on the four short-term rates, namely the 1-month and the 3-month eurodollar rate, the fed fund rate and the 3-month t-bill rate. We consider two time spans: the 1976-2002 period that was examined earlier (Table 3 and 4) and the shorter period between 1983 and 2002 (Table 5 and 6) to avoid inclusion of the high volatility/high rate phenomena of the years between 1980 and 1983. Over the longer time period the level test implies failure to reject the null of nonstationarity except for a number of factors equal to one (c.f. Table 3). Over the shorter period between 1980 and 1983 we fail to reject only in association with four factors (c.f. Table 5). Nonetheless, four is the number of factors that both the BIC and IC criterion choose for the level test in both cases. The demeaned test continues to largely support nonstationarity in both cases.

Based on the reported tests, it is sensible to doubt about the stationarity properties of the short end of the term structure. On the other hand, stationarity cannot be excluded a-priori. In what follows, coherently with a large number of recent papers, we model the short-term interest rate process as a scalar diffusion (the solution to (14) above, that is) and estimate drift and diffusion using (48) and (49) under the assumption of recurrence as in Bandi (1998). We apply the methodology to the federal fund rates between 1976 and 2002. Fig. 1 contains the series in level. In Fig. 2 we graph the series in first differences. The information that we obtain from both figures is standard and characterizes virtually any short-term interest rate series (c.f. Aït-Sahalia (1996), for example). In particular, noticeable are the high rate/high volatility features of the time period between 1980 and 1982.

In Fig. 3 and 4 we plot the estimated drift and diffusion function for levels that are often visited, that is for levels between 3% and 16%, using four smoothing parameters, i.e. 2%, 2.2%, 2.4% and 3%. The drift appears to be mean-reverting at both ends of the examined range, while the diffusion behaves as a function proportional to r^θ, where r is the interest rate level and θ is a constant between 2 and 3. This class of functions is typically called 'constant elasticity of variance' (CEV, henceforth) following Cox (1975).

A closer look at the previous results requires a more formal assessment of statistical

[10]We thank the Federal Reserve Bank of Chicago for providing the data.

accuracy through the use of standard errors. In Fig. 5 and 6 we plot the nonparametric drift and diffusion and corresponding confidence intervals for bandwidth choices equal to 2.4% and 2% respectively, over the period 1976-2002 and, for comparison, over the shorter period 1983-2002. Several observations are in order. The substantial reversion to the mean induced by the drift at high rates is not statistically significant (c.f. Bandi (1998)). In particular, it is hard to reject the assumption of a zero drift in the restricted range between 3% and 16%. The CEV features of the diffusion function are estimated accurately (c.f. Conley *et al.* (1997) for a similar result in the presence of the same time series examined over the shorter period from January 2, 1970 to January 29, 1997).

When taken face value the above results suggest nonstationary martingale behaviour for the short-term rate, thus confirming (on an ex-post basis) the necessity of being cautious about the statistical assumptions that are required to assist statistical inference on the short-end of the term structure. On the other hand, as discussed by Bandi (1998) in the context of eurodollar rates, a richer specification than that implied by the reported nonparametric estimates appears to be necessary to model the process over its entire admissible range. In effect, the natural scale diffusion that solves $dr_t = cr_t^{\theta} dB_t$ with $c > 0$ and $\theta \in (2, 3)$ behaves as a nonstationary (being the speed measure not integrable) martingale with an attracting boundary at 0 since the scale function at 0 is finite (c.f. Karlin and Taylor (1981)). In spite of the fact that the 0 boundary is attainable in finite time with probability zero, it is sensible to believe that more reversion to the center of the admissible range (l, u) is necessary for modelling purposes (c.f. Bandi (1998)). In fact, the above scale diffusion is not recurrent over $(0, \infty)$.

It should be noted that, even if we believed in (the economically unreasonable assumption of) transience, the estimated diffusion function would be consistent since, as pointed out earlier, the infinitesimal second moment can be identified over a fixed span of data and, as a consequence, does not necessitate infinite visits to a level for pointwise consistent estimation. As for the drift, its theoretical consistency would not be guaranteed but our previous discussion suggests that, in the presence of transient data, informative estimates can still be obtained at spatial levels that are visited often by the sampled process, that is at levels whose associated local time is large (c.f. Bandi (1998)).

A more complicated specification for the diffusion can induce recurrence and resolve the absorption problem at low rates. If $\sigma^2(r) = c_0 + c_1|r|^{\theta}$ with $c_0 > 0$, $c_1 \geq 0$ and $\theta \in (2, 3)$, then the resulting natural scale diffusion is positive-recurrent (c.f. Example 1 in Section 3) but the admissible range becomes $(-\infty, \infty)$. Such a range is obviously not acceptable as far as nominal rates are concerned. Alternatively, reversion to the center of the range of the process can be induced by a sufficiently positive drift at low rates. It is not clear that such drift is in the data. In Fig. 7 we report functional estimates of the measure of pull to the right that was put forward by Conley *et al.* (1997), i.e. $\frac{\mu(.)}{2\sigma^2(.)}$.[11] While the point estimates

[11]Given a point $x \in (l, u)$, it is easy to show that the probability p of reaching $x + \varepsilon$ prior to $x - \varepsilon$ is given

suggest some pull to the center of the distribution, mostly for small interest rates, such effect does not seem to be statistically significant. As observed earlier, the series appears to have a dynamic evolution that is close to that of a continuous (local) martingale over its empirical range.

To conclude, the limitations of scalar diffusions as modelling devices intended to capture the global properties of interest rate data suggest that more involved specifications might and should be entertained. For example, allowing for jumps (as in Section 4) would permit to better capture the higher moment properties of the data (c.f. Das (2002) for a parametric study of jumps in federal fund rates and Johannes (1999) for an application of the nonparametric estimation procedure in Section 4 to t-bill rates). The use of multivariate specifications (as in Section 5) would respond to the well-known empirical necessity of allowing for more than one underlying factor when describing the temporal evolution of the entire term structure of interest rates (c.f. Litterman and Scheinkman (1991)).

7 Concluding Remarks

In surveying the tools that have been recently introduced to describe and study the formulation and estimation of classes of potentially nonstationary continuous-time Markov models, this Chapter illustrates the important role that is played by the assumption of recurrence in these estimation problems. The focus of our discussion has been functional continuous-time specifications for asset pricing. But similar arguments in favor of minimal conditions on the underlying structure of the process of interest can be put forward when dealing with parametric models and discrete-time series. Sometimes empirical researchers may be a lot more comfortable avoiding restrictions like stationarity or arbitrary mixing conditions on the processes they are modeling. In the same circumstances, it might also seem inappropriate to impose explicit nonstationary behaviour in the specification. Indeed, many practical situations arise where neither stationarity nor nonstationarity can be safely ruled out in advance and, in such situations, the assumption of recurrence appears to be a suitable alternative condition that permits a wide range of plausible sample behaviors and includes both stationary and nonstationary processes. Interestingly, statistical inference can often be carried out in recurrent models using limiting laws defined in terms of random norming (the averaged kernel in the definition of the estimated local time being an example, c.f. (51) and (54) for instance). Such random norming captures the divergence features of time series with various degrees of recurrence and allows the user to be agnostic about the recurrence features of the processes of interest. The practical advantage of this fact is

---

by

$$p_x(\varepsilon) = \frac{1}{2} + \frac{\mu(x)}{2\sigma^2(x)}\varepsilon + o(\varepsilon),$$

(c.f. Conley *et al.* (1997)).

apparent. While standard asymptotic theory treats stationary and nonstationary models differently in deriving implications for statistical inference, reliance on recurrence permits one to consider both cases as subcases of a more general theory of inference. Additionally, even when the existance of a stationary density appears to be an unquestionable feature of the data and/or is dictated by economic theory, the dynamic structure of Markov processes renders conventional forms of mixing not crucial to derive limiting results and, consequently, vital tools for statistical analysis.

Having made these points, we should add the qualification that the use of recurrence as an identifying condition is still in its infancy in the econometrics literature. In 1989 Yakovitz conjectured that "...in the Markov case the mixing assumptions are not essential...Even in the absence of a stationary distribution, under conditions general enough to include unbounded random walks and ARMA processes, [nonlinear] regression estimation is possible. We require only stationarity of the transition law, not of the process." Harris recurrence is the identifying assumption in Yakovitz's work, but the treatment in that paper only focuses on the discrete-time ergodic case. More recently, kernel density estimation for real-valued positive Harris recurrent Markov Chains is discussed in Athreya and Atuncar (1998). Phillips and Park (1998) study the nonparametric estimation of nonstationary time series embeddable in Brownian motion (which is known to be $\frac{1}{2}$-null recurrent). Karlsen *et al.* (1999) focus on the functional estimation of cointegrating relations between β-null recurrent discrete-time Markov chains (see, also, Karlsen and Tjøstheim (1999)). Closer to the methods discussed in this Chapter, where neither stationarity nor nonstationarity of the null-recurrent type are ever imposed explicitly, is the approach in Moloche (2001). Moloche (2001) tackles the nonparametric estimation of (potentially cointegrating) regressions between (either null or positive) Harris recurrent discrete-time Markov sequences.

The methods reviewed in the present Chapter, along with the recent treatments mentioned above, have helped to lay some foundations for econometric inference with continuous and discrete-time series under mild assumptions on their parametric form and statistical evolution. But the field is a new one and, as this Chapter has suggested, there is much more to be done.

References

[1] AÏT-SAHALIA, Y. (1996): "Testing Continuous-Time Models of the Spot Interest Rate," *The Review of Financial Studies*, 2, 385–426.

[2] AÏT-SAHALIA, Y., L.P. HANSEN and J. SCHEINKMAN (2001): "Operator Methods for Continuous-Time Markov Processes," *This volume.*

[3] ANDREWS, D.W.K. (1989): "Asymptotics for Semiparametric Econometric Models: III. Testing and Examples," *C.F.D.P. No. 910, Yale University.*

[4] ATHREYA, K.B. and G.S. ATUNCAR (1998): "Kernel Estimation for Real-Valued Markov Chains," *Sankhyā,* 60, 1-17.

[5] AZÉMA, J., M. KAPLAN-DUFLO and D. REVUZ (1967): "Mesure Invariante Sur Les Classes Récurrentes des Processus de Markov," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete,* 8, 157–181.

[6] BAI, J. and S. NG (2001): "Determining the Number of Factors in Approximate Factor Models," *Econometrica,* 70, 191-223.

[7] BANDI, F. (1998): "Short-Term Interest Rate Dynamics: A Spatial Approach," *Journal of Financial Economics, Forthcoming.*

[8] BANDI, F. and G. MOLOCHE (2001): "On the Functional Estimation of Multivariate Diffusion Processes," *Unpublished paper, The University of Chicago and M.I.T.*

[9] BANDI, F. and T. NGUYEN (1999): "Fully Nonparametric Estimators for Diffusions: a Small Sample Analysis," *Unpublished paper, The University of Chicago.*

[10] BANDI, F. and T. NGUYEN (2000): "On the Functional Estimation of Jump-Diffusion Processes," *Journal of Econometrics, Forthcoming.*

[11] BANDI, F. and P. C. B. PHILLIPS (1998): "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, *Forthcoming.*

[12] BANDI, F. and P. C. B. PHILLIPS (1999): "Semiparametric Methods for Diffusion Model Estimation," *Unpublished paper, The University of Chicago and Yale University.*

[13] BIBBY, B. M., M. JACOBSEN and M. SØRENSEN (2001): "Estimating functions for diffusions," *This volume.*

[14] BOSQ, D. (1997): "Temps Local et Estimation Sans Biais de la Densite' en Temps Continu," C.R. Acad. Sci. Paris t. 325, 1, 527-530.

[15] BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes.* Second Edition. Springer-Verlag: New York.

[16] BOSQ, D. and Y. DAVYDOV (1998): "Local Time and Density Estimation in Continuous Time." Publ. IRMA Univ. Lille 1, Vol. 44, III.

[17] BRUGIÈRE, P. (1991): "Estimation de la Variance d'un Processus de Diffusion dans le Cas Multidimensionnel," *C. R. Acad. Sci,* T. 312, Serie I, 13, 999-1005.

[18] BRUGIÈRE, P. (1993): "Théorém de Limite Centrale Pour un Estimateur Non Paramétrique de la Variance d'un Processus de Diffusion Multidimensionnelle," *Ann. Inst. Henri Poincaré,* Vol. 29, n. 3, 357-389.

[19] CAMPBELL, J.Y., A. LO and A.C. MACKINLAY (1997): *The Econometrics of Financial Markets.* Princeton University Press. Princeton, New Jersey.

[20] CHEN, X. (1999): "How Often Does a Harris Recurrent Markov Chain Recur?," *The Annals of Probability,* 3, 1324-1346.

[21] CHEN, X. (2000): "On the Limit Laws of the Second Order for Additive Functional of Harris Recurrent Markov Chains," *Probab. Theory Relat. Fields,* 116, 89-123.

[22] CHEN, X., L.P. HANSEN, and M. CARRASCO (1999): "Nonlinearity and Temporal Dependence," *Unpublished paper.*

[23] COX, J.C. (1975): *Notes on Option Pricing I: Constant Elasticity of Variance Diffusion.* Lecture notes. Stanford University.

[24] COX, J.C., J.E. INGERSOLL and S.A. ROSS (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica,* 53, 385-408.

[25] DARLING, D. A. and M. KAC (1957): "On Occupation Times for Markoff Processes," *Transactions of the American Mathematical Society*, 84, 444–458.

[26] DAS, S. R. (2002): "The Surprise Element: Jumps in Interest Rates," *The Journal of Econometrics,* 106, 27-65.

[27] DUFFIE, D. (1996): *Dynamic Asset Pricing Theory.* Princeton University Press: Princeton.

[28] DURRETT, R. (1996): *Stochastic Calculus. A Practical Introduction.* CRC Press: New York.

[29] ETHIER, S.N. and T.G. KURTZ (1986): *Markov Processes: Characterization and Convergence.* New York: John Wiley & Sons.

[30] FAMA, E., and G. W. SCHWERT (1977): "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115-146.

[31] FAN, J. (1992): "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998-1004.

[32] FAN, J. (1993): "Local Linear Regression Smoothers and their Minimax Efficiencies," *The Annals of Statistics*, 21, 196-216.

[33] FAN, J. and I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications.* Chapman and Hall.

[34] FLORENS-ZMIROU, D. (1993): "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30, 790-804.

[35] GALLANT, R. and G. TAUCHEN (2001): "Simulated Methods and Indirect Inference for Continuous-Time Models," *This volume.*

[36] GEMAN, D. and J. HOROWITZ (1978): "Occupation Densities," *The Annals of Probability*, 1, 1-67.

[37] GEMAN, S.A. (1979): "On a Common Sense Estimator for the Drift of a Diffusion," *Working Paper, Brown University.*

[38] GIKHMAN, I.I. and A.V. SKOROHOD (1972): *Stochastic Differential Equations.* Springer-Verlag: New York.

[39] GOURIEROUX, C. and J. JASIAK (2001): *Financial Econometrics: Problems, Models and Methods.* Princeton University Press. Princeton.

[40] HANSEN, L.P. and T.J. SARGENT (1983): "The Dimensionality of the Aliasing Problem in Models with Rational Spectral Densities," *Econometrica,* 50, 377-387.

[41] HANSEN, L.P. and J. A. SCHEINKMAN (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica,* 63, 767-804.

[42] HÄRDLE, W. (1990): *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

[43] HÄRDLE, W. and O. LINTON (1994): "Applied Nonparametric Methods," in D. McFadden and R.F. Engle (eds.), *The Handbook of Econometrics,* Vol. IV, New York, North Holland, 2295-2339.

[44] HÖPFNER, R. and E. LÖCHERBACH (2001): "Limit Theorems for Null-Recurrent Markov Processes," *Preprint N. 22, Fachbereich Mathematik, Universität Mainz.*

[45] HÖPFNER, R. and Y. KUTOYANTS (2001): "On a Problem of Statistical Inference in Null Recurrent Diffusions," *Preprint N. 7, Fachbereich Mathematik, Universität Mainz.*

[46] JACOD, J. (1997): "Nonparametric Kernel Estimation of the Diffusion Coefficient of a Diffusion," *Prépublication No. 405. du Laboratoire de Probabilités de l'Université Paris VI.*

[47] JACOD, J. (2001): "Inference for Stochastic Processes," *This volume.*

[48] JIANG, G.J. and J. KNIGHT (1997): "A Nonparametric Approach to the Estimation of Diffusion Processes, with An Application to a Short-Term Interest Rate Model," *Econometric Theory,* 13, 615-645.

[49] JOHANNES, M. (1999): "Jumps in Interest Rates: a Nonparametric Approach," *Working paper, Columbia Business School.*

[50] JOHANNES, M. and N. POLSON (2001): "Numerical Bayesian Methods for Estimating Continuous-Time Models," *This volume.*

[51] KARATZAS, I. and S. E. SHREVE (1988): *Brownian Motion and Stochastic Calculus.* Springer-Verlag: New York.

[52] KARLIN, S. and H. M. TAYLOR (1981): *A Second Course in Stochastic Processes.* Academic Press: New York.

[53] KARSEN, H. A. and D. TJØSTHEIM (1998): "Nonparametric Estimation in Null Recurrent Time Series Nonlinear Generalization," *Sonderforschungsbereich*, 373, Humboldt-Universität zu Berlin.

[54] KARSEN, H. A., T. MYKLEBUST and D. TJØSTHEIM (1999): "Nonparametric Estimation in a Nonlinear Cointegration Type Model," *Working paper*, *The University of Bergen.*

[55] KASPI, H. and A. MANDEBAUM (1994): "On Harris Recurrence in Continuous-Time," *Mathematics of Operation Research*, 19, 211–222.

[56] KHASMINSKII, R. Z.(1980): *Stochastic Stability of Differential Equations.* Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands.

[57] KHASMINSKII, R. Z.(2001): "Limit Distributions of Some Integral Functionals for Null-Recurrent Diffusions," Stochastic Processes and their Applications, 92, 1-9.

[58] KHASMINSKII, R. Z. and G. YIN (2000): "Asymptotic Behavior of Parabolic Equations Arising from One-Dimensional Null-Recurrent Diffusions," Journal of Differential Equations, 161, 154-173.

[59] LITTERMAN, R. and J. SCHEINKMAN (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income,* 1, 54-61.

[60] MENALDI, J.L. and M. ROBIN (1999): "Invariant Measure for Diffusions with Jumps," *Applied Mathematics and Optimization,* 40, 105-140.

[61] MEYN, S.P. and R.L. TWEEDIE (1993): *Markov Chains and Stochastic Stability.* Springer-Verlag. London (1993).

[62] MERTON, R.C. (1990): *Continuous-Time Finance.* Blackwell Publishers.

[63] MOLOCHE, G. (2000): "Local Nonparametric Estimation of Scalar Diffusions," *Unpublished paper*, MIT.

[64] MOLOCHE, G. (2001): "Kernel Regression for Non-Stationary Harris Recurrent Processes," *Unpublished paper*, MIT.

[65] MOON, H.R. and B. PERRON (2001): "Testing for a Unit Root in Panels with Dynamic Factors," *Unpublished paper,* USC and University of Montreal.

[66] PAGAN, A. and A. ULLAH (1999): *Nonparametric Statistics.* Cambridge University Press.

[67] PIAZZESI. M. (2001): "Affine Term-Structure Models," *This volume.*

[68] PHILLIPS, P.C.B. (1973): "The Problem of Identification in Finite Parameter Continuous-Time Models," *The Journal of Econometrics,* 4, 351-362.

[69] PHILLIPS, P.C.B. (1998): "Econometric Analysis of Fisher's Equation," *Cowles Foundation Discussion Paper, No. 1180. Presented at the Irving Fisher Conference, Yale University, 1998.*

[70] PHILLIPS, P.C.B. (2001): "Descriptive Econometrics for Nonstationary Time Series with Empirical Illustrations," *Journal of Applied Econometrics,* 16m, 389-413.

[71] PHILLIPS, P.C.B. AND J. PARK (1998): "Nonstationary Density Estimation and Kernel Autoregression," *Cowles Foundation Discussion Paper, No. 1181, Yale University.*

[72] PHILLIPS, P.C.B. and D. SUL (2001). "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence" *Cowles Foundation Discussion Paper, No. 1362, Yale University.*

[73] POLLACK, M. and D. SIEGMUND (1985): "A Diffusion Process and its Applications to Detecting a Change in the Drift of Brownian Motion," *Biometrica,* 72, 267-280.

[74] PRAKASA-RAO, B.L.S. (1983): *Nonparametric Functional Estimation.* New York, Academic Press.

[75] PROTTER, P. (1990): *Stochastic Integration and Differential Equations.* Springer-Verlag: New York.

[76] RAY, D. (1963): "Sojourn Times of Diffusion Processes," *Illinois J. of Math.,* 7, 425-493.

[77] REVUZ, D. AND M. YOR (1998): *Continuous Martingales and Brownian Motion.* Third Edition, Springer-Verlag: New York.

[78] SILVERMAN, B (1986): *Density Estimation for Statistics and Data Analysis.* London, Chapman and Hall.

[79] STONE, D. W. (1992): *Multivariate Density Estimation. Theory, Practice and Visualization.* Wiley Series in Probability and Mathematical Statistics.

[80] SUNDARESAN, S. M. (2000): "Continuous-Time Methods in Finance: A Review and an Assessment," *Journal of Finance*, 55, 1569-1622.

[81] TOROUS W., R. VALKANOV and S. YAN (2001): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," *Unpublished working paper.*

[82] VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics,* 5, 177-188.

[83] YAKOWITZ, S. (1989): "Nonparametric Density and Regression Estimation for Markov Sequences Without Mixing Assumptions," *Journal of Multivariate Analysis,* 30, 124-136.

[84] YOR, M. (1978): "Sur la Continuité des Temps Locaux Associés à Certaines Semi-Martingales," Astérisque 52-53 (1978) 219-221.

[85] YOR, M. (1983): "Le Drap Brownien Comme Limite en Loi de Temps Locaux Lineaires," Sém. Prob. XVII. Lecture Notes in Mathematics, vol. 986. Springer, Berlin Heidelberg New York, 89-105.

Estimated # of factors
$IC = 8$
$BIC = 1$

# of factors	T-stat	ρ
1	-2.865	0.99982
2	-2.670	0.99984
3	-3.617	0.99945
4	-2.472	0.99962
5	-3.688	0.99952
6	-2.954	0.99961
7	-5.465	0.99862
8	-6.018	0.99826

Table 1. The Moon and Perron (2001) level test for nonstationarity.
We apply it to 1-month, 3-month and 6-month eurodollar rates, the federal fund rate, 3-month and 6-month t-bill rates, and constant maturity rates between 1 and 10 years (namely, 1, 2, 3, 5, 7 and 10 years). We use 6,418 observations from June 1, 1976 to March 15, 2002.

Estimated # of factors
$IC = 8$
$BIC = 1$

# of factors	T-stat	ρ
1	1.906	1.00016
2	3.047	1.00026
3	2.385	1.00024
4	2.912	1.00034
5	6.631	1.00060
6	5.997	1.00061
7	20.561	1.00216
8	23.278	1.00249

Table 2. The Moon and Perron (2001) demeaned test for nonstationarity.
We apply it to 1-month, 3-month and 6-month eurodollar rates, the federal fund rate, 3-month and 6-month t-bill rates, and constant maturity rates between 1 and 10 years (namely, 1, 2, 3, 5, 7 and 10 years). We use 6,418 observations from June 1, 1976 to March 15, 2002.

Estimated # of factors
$IC = 4$
$BIC = 4$

# of factors	T-stat	ρ
1	-1.971	0.99969
2	-1.516	0.99974
3	-1.612	0.99955
4	-1.507	0.99946

Table 3. The Moon and Perron (2001) level test for nonstationarity.
We apply it to 1-month and 3-month eurodollar rates, the federal fund rate and 3-month t-bill rates.
We use 6,418 observations from June 1, 1976 to March 15, 2002.

Estimated # of factors
$IC = 4$
$BIC = 4$

# of factors	T-stat	ρ
1	2.141	1.00031
2	2.644	1.00048
3	3.932	1.00082
4	6.824	1.00154

Table 4. The Moon and Perron (2001) demeaned test for nonstationarity.
We apply it to 1-month and 3-month eurodollar rates, the federal fund rate and 3-month t-bill rates.
We use 6,418 observations from June 1, 1976 to March 15, 2002.

Estimated # of factors
$IC = 4$
$BIC = 4$

# of factors	T-stat	ρ
1	-2.448	0.99975
2	-1.765	0.99973
3	-3.971	0.99682
4	-1.566	0.99925

Table 5. The Moon and Perron (2001) level test for nonstationarity.
We apply it to 1-month and 3-month eurodollar rates, the federal fund rate and 3-month t-bill rates.
We use 4,796 observations from January 1, 1983 to March 15, 2002.

Estimated # of factors
$IC = 4$
$BIC = 4$

# of factors	T-stat	ρ
1	-0.253	0.99995
2	0.779	1.00020
3	0.212	1.00005
4	8.852	1.00313

Table 6. The Moon and Perron (2001) demeaned test for nonstationarity.
We apply it to 1-month and 3-month eurodollar rates, the federal fund rate and 3-month t-bill rates.
We use 4,796 observations from January 1, 1983 to March 15, 2002.

Fig.1 : Federal fund rate series from 1976 to 2002

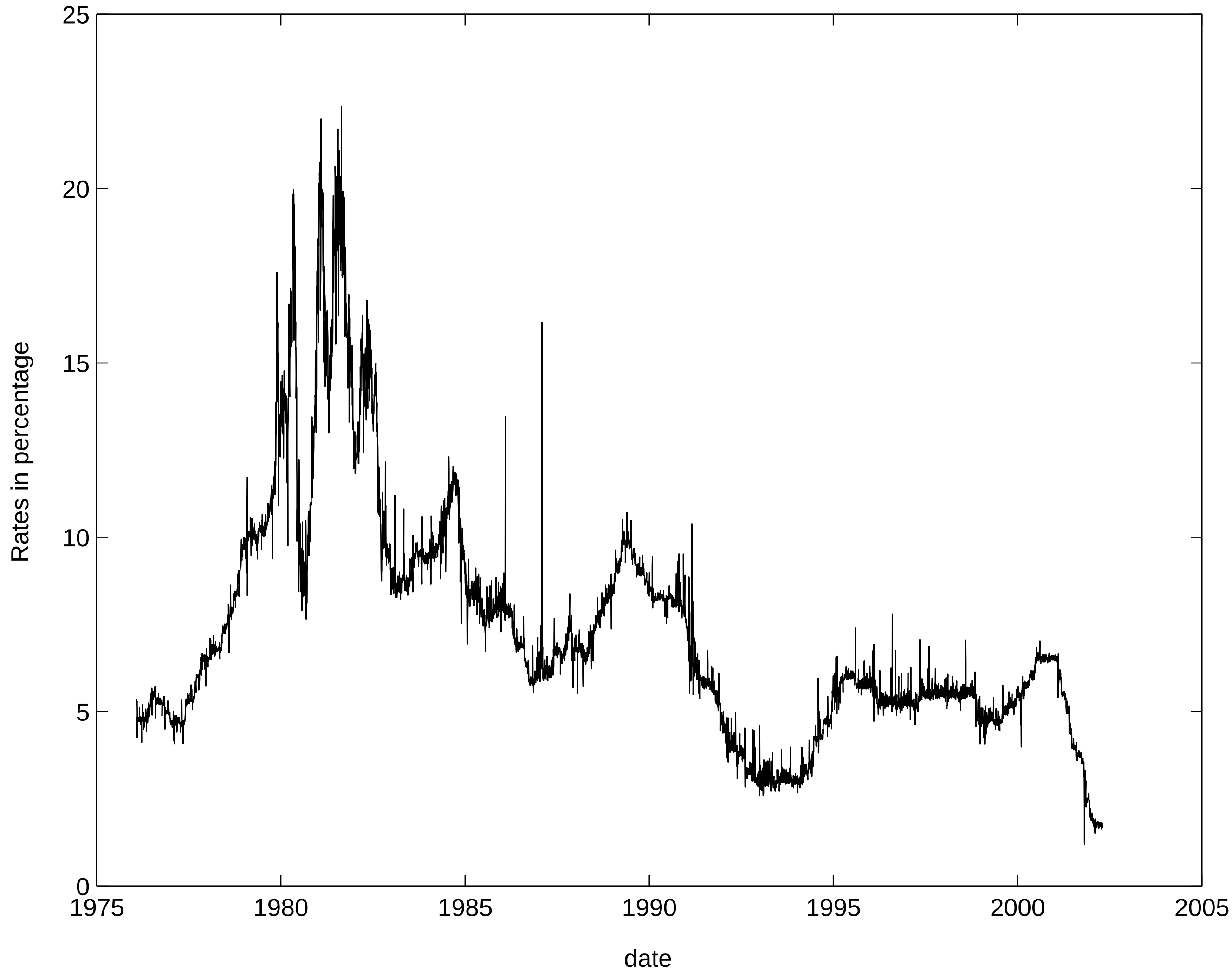
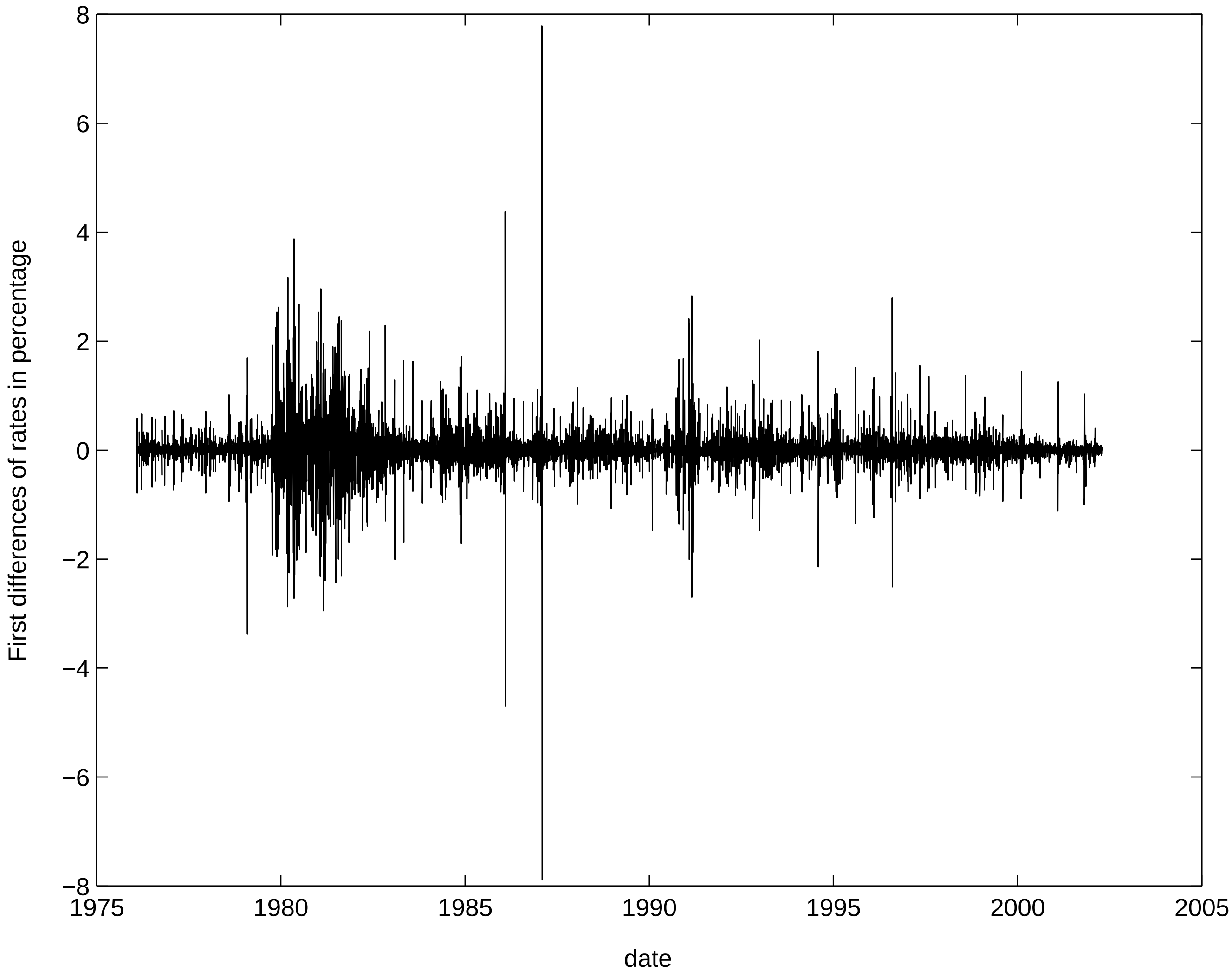


Fig.2 : Federal fund rate series from 1976 to 2002



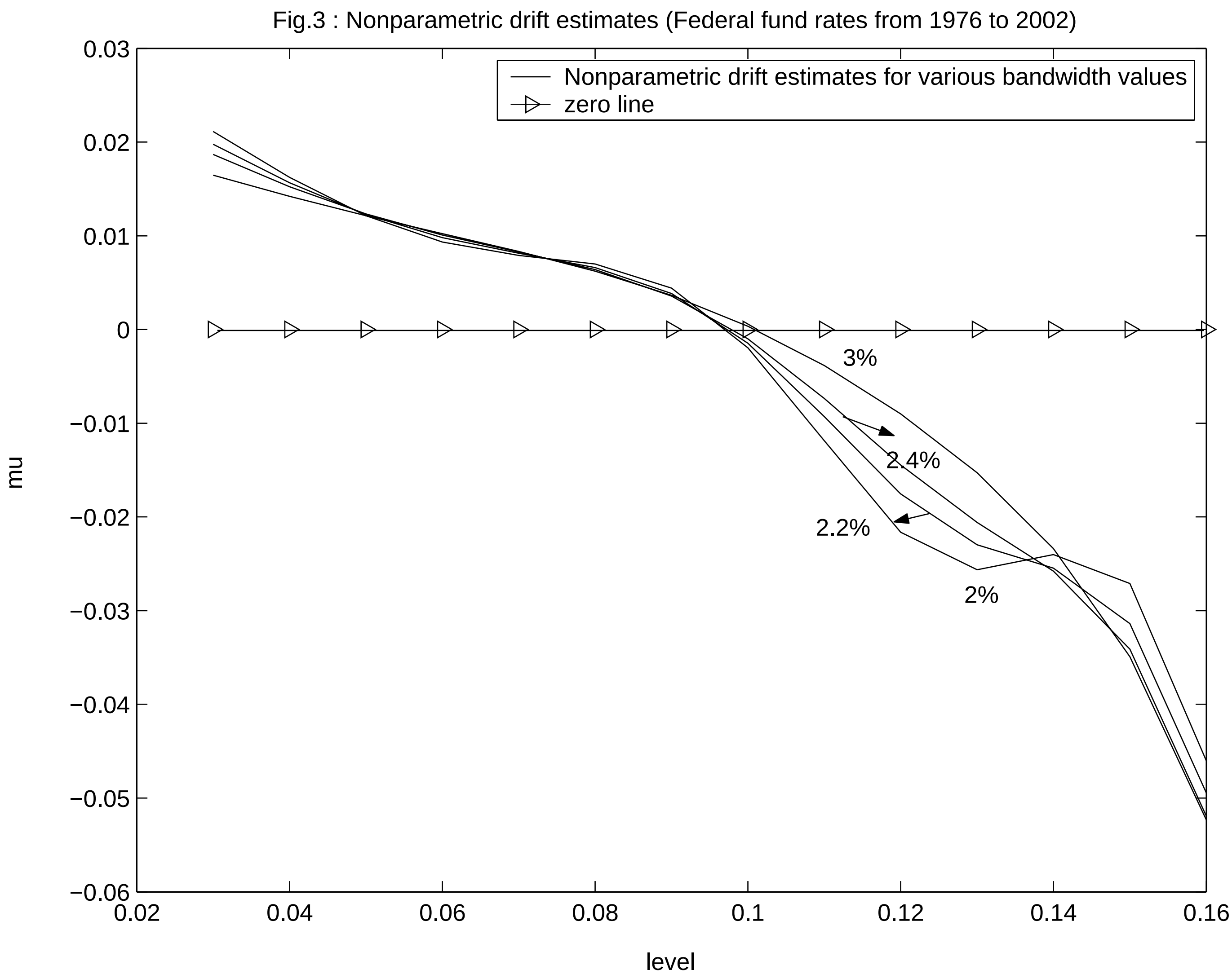


Fig.3 : Nonparametric drift estimates (Federal fund rates from 1976 to 2002)
Nonparametric drift estimates for various bandwidth values
zero line
3%
2.4%
2.2%
2%
mu
level
0.03
0.02
0.01
0
−0.01
−0.02
−0.03
−0.04
−0.05
−0.06
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16

Fig.4 : Nonparametric diffusion estimates (Federal fund rates from 1976 to 2002)

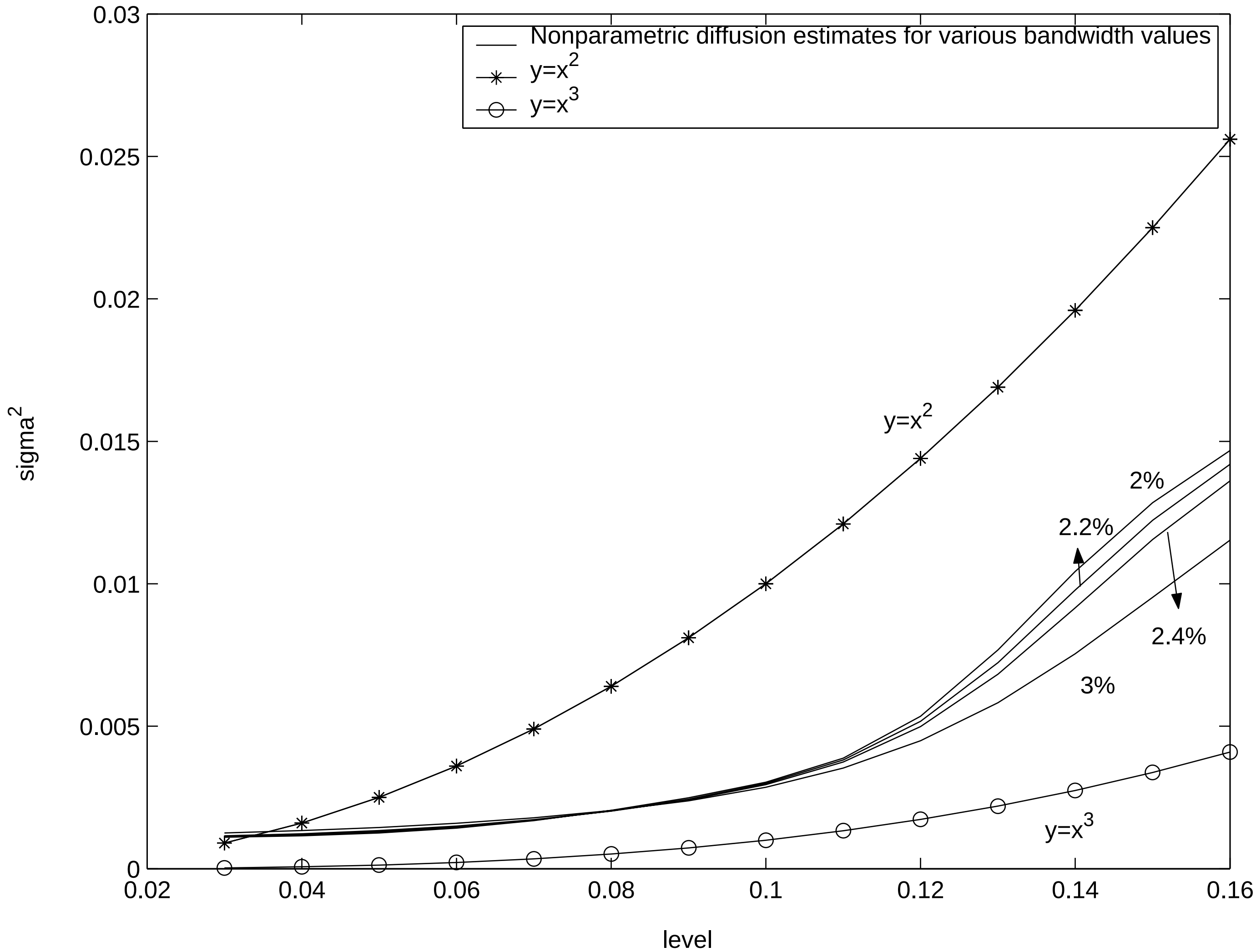


Fig.5 : Nonparametric drift estimates (Federal fund rates 1976−2002 and 1983−2002)

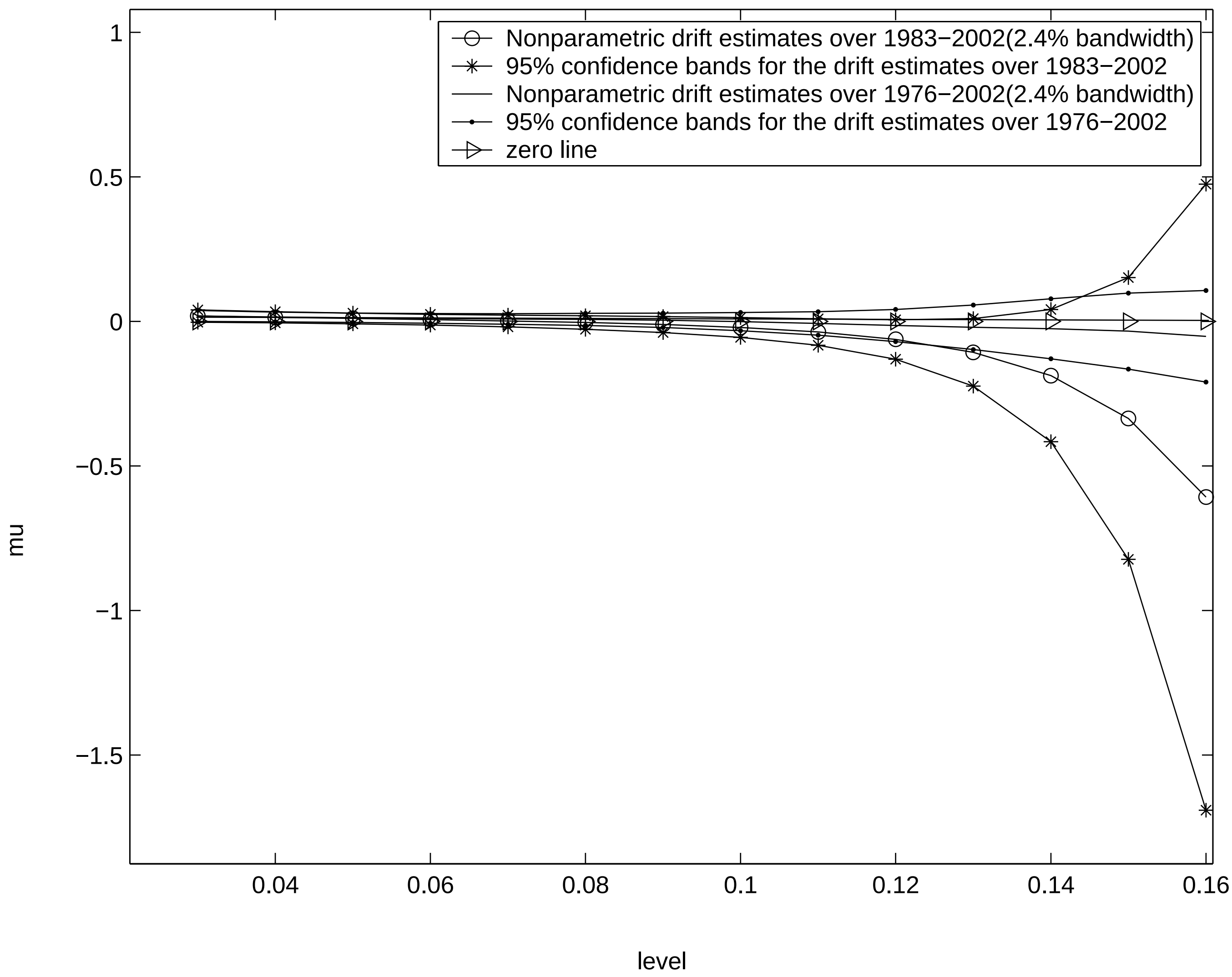


Fig.6 : Nonparametric diffusion estimates (Federal fund rates 1976−2002 and 1983−2002)

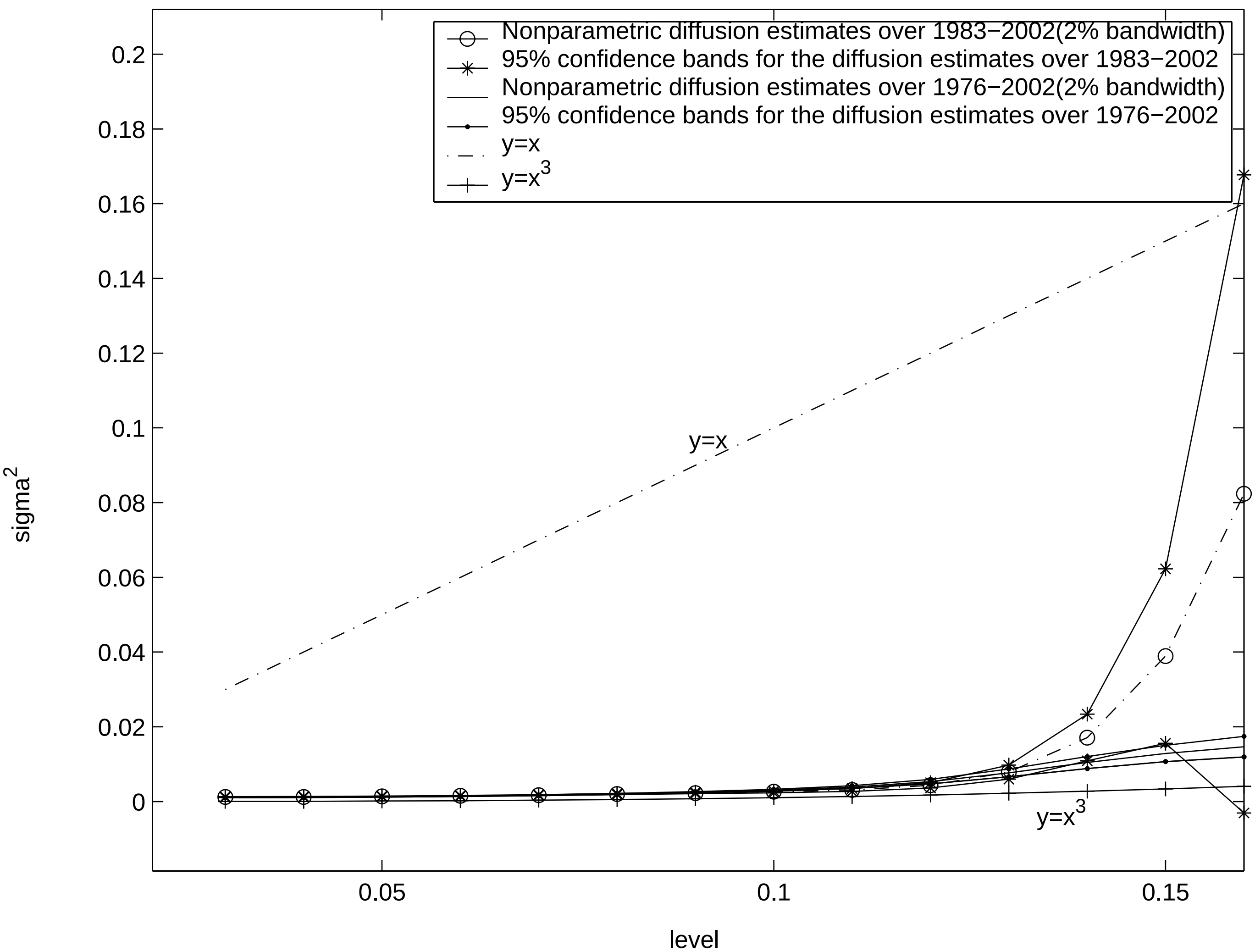


Fig.7 : Nonparametric pull measure (Federal fund rates 1976−2002 and 1983−2002)